THE TWENTY-FIRST YEARBOOK

OF THE

NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

INTELLIGENCE TESTS AND THEIR USE

PART I

THE NATURE, HISTORY, AND GENERAL PRINCIPLES OF INTELLIGENCE TESTING

PART II

THE ADMINISTRATIVE USE OF INTELLIGENCE TESTS

Prepared by the Society's Committee and Edited by

GUY MONTROSE WHIPPLE

This Yearbook will be Discussed at the Chicago Meeting of the National Society, Saturday, February 25, and Tuesday, February 28, 1922. 8:00 p. m.

PUBLIC SCHOOL PUBLISHING COMPANY
BLOOMINGTON, ILLINOIS
1922

OFFICERS OF THE SOCIETY
for the Chicago Meeting

President

FREDERICK JAMES KELLY
University of Kansas, Lawrence, Kansas

Vice-President

LIDA LEE TALL
State Normal School, Towson, Maryland

Secretary-Treasurer

GUY MONTROSE WHIPPLE
University of Michigan, Ann Arbor, Michigan

Executive Committee

(The Year Indicates Date of Expiration of Term)

JOSEPH C. BROWN (1922)
State Normal School, St. Cloud, Minnesota

PAUL WHITFIELD HORN (1923)
Superintendent of Schools, Houston, Texas

STEPHEN S. COLVIN (1924)
Brown University, Providence, Rhode Island

HENRY W. HOLMES (1925)
Harvard University, Cambridge, Massachusetts

Board of Trustees

W. W. KEMP (1922)
State Normal School, San Jose, California

FRANK W. BALLOU (1923)
Superintendent of Schools, Washington, D. C.

W. W. CHARTERS (1924)
Carnegie Institute of Technology, Pittsburgh, Pennsylvania

TABLE OF CONTENTS

EDITOR'S PREFACE

At the Atlantic City meeting in February, 1921, the Commission of the National Education Association on Co-ordination of Research Agencies, passed a resolution and appointed a committee to ask the National Society for the Study of Education to devote one of its *Yearbooks* to the discussion of intelligence tests. A similar action was taken at the same time by the National Association of Directors of Educational Research, and Messrs. B. R. Buckingham and George Melcher conveyed to the Executive Committee of this Society the attitude of the two Associations just mentioned. It so happened that at the same time the Executive Committee of this Society were considering a *Yearbook* dealing with intelligence testing, so that its decision to produce such a *Yearbook* represents the desires of all three associations. Professor Stephen S. Colvin was formally appointed chairman of a special Committee to solicit contributions and assemble the material for the 1922 *Yearbook*, with the understanding that emphasis should be laid upon group intelligence testing and particularly upon the administrative aspects of this important educational development. The present *Yearbook*, therefore, represents the labors of the Committee headed by Professor Colvin, and is presented as a contribution by the National Society for the Study of Education on the theme proposed by its own Executive Committee, by the National Education Association's Commission, and by the National Association of Directors of Educational Research. The editor is responsible for the final revision of the material.

GUY M. WHIPPLE.

INTRODUCTION

The most significant and important movement in the field of education during the past decade has been the rapid development and the constantly increasing use of scientific measurements. These in the main have been of two sorts—measurements to ascertain the native ability of the pupil, and measurements to determine his school attainment. The first of these has to do with so-called "intelligence tests," or "mentality tests," and the second with tests for specific school subjects. Intelligence tests were first systematically undertaken by Binet more than fifteen years ago, but it is only within more recent years that these tests and others of an analogous nature have been extensively employed in school practice.

In 1897 Dr. J. M. Rice published in the *Forum* two articles giving an account of his investigations of the spelling abilities of school children in the United States. The simple tests that he employed were the first definite attempt made on an extensive scale to measure any aspect of school achievement. For this reason Dr. Rice has been called the "father of educational measurements." Since this early attempt, the movement to measure school attainments in a fundamental and scientific way has grown to astonishing proportions.

The growth and practical application of intelligence tests has paralleled that of tests to measure school products. The two movements have gone hand in hand, as indeed, they should. Both must be used in conjunction if we wish to know the real facts about actual achievement of pupils and the efficiency of a teacher, a room, a building, or a school system.

The recent wide acceptance of these two agencies for determining school achievement has been on the whole decidedly beneficial. However, the character of tests and their theoretical and practical values have been misunderstood in part, and the result too often has been either an unreasoning and blind antagonism or a superlative and uncritical acceptance of these means for discovering and directing pupils' abilities and attainments.

To those who believe in the fundamental value of educational testing, the antagonism of some of its opponents has been annoying, while the unrestrained enthusiasm of some of its uncritical supporters has been alarming. It is in the field of mental testing that the greater danger resides, since here the nature, objects, and practical values of testing are more easily misunderstood than in the field of the measurement of educational products.

For the purposes of correcting some of these errors and misunderstandings and of explaining in a clear and accurate manner the theory, nature, and practical use of intelligence tests, the present *Yearbook* has been compiled. It is composed of two parts. In Part I the more theoretical, general, and technical aspects of mental testing are set forth in such a manner, it is hoped, that the treatment may be easily understood by those who have little expert knowledge of, or skill in, the matters here considered. Indeed, it is the aim in this part of the *Yearbook*, as well as in the following part, to set forth the facts in regard to mental testing in as simple and direct a way as possible, so that all who are interested in the subject may get a real insight into the theory and the uses of mental testing.

Part I attempts to show just what is to be understood by the term "general intelligence," to indicate how this may be measured and to show the steps by which mental tests have grown up and some of their most essential characteristics. Further, the attempt is made to acquaint the teacher and administrator with the correct methods of studying and evaluating the results of mental testing. A descriptive bibliography is added which furnishes information in regard to the various group tests of intelligence now available. A brief chapter is added on the importance of measurement in education generally.

Part II takes up in some detail the administrative uses of intelligence tests in various grades of instruction, beginning with the primary grades and ending with the college and university. In the discussions in this part of the book the purpose is to set forth in some detail the procedure and results of mental testing as far as they relate to matters of instruction and administration.

The Committee hopes that the *Yearbook* will prove its worth as a guide to those who wish to understand the significance of mental tests and who seek to employ them for the betterment of the schooi product. If this hope is to any extent realized, the Committee feels that its labors will not have proved in vain.

HELEN DAVIS,	AGNES L. ROGERS,
BESSIE LEE GAMBRILL,	HAROLD O. RUGG,
HENRY W. HOLMES,	M. R. TRABUE,
WARREN K. LAYTON,	E. L. THORNDIKE,
W. S. MILLER,	G. M. WHIPPLE,
RUDOLPH PINTNER,	STEPHEN S. COLVIN, *Chairman.*

CHAPTER I

MEASUREMENT IN EDUCATION

E. L. Thorndike
Professor of Educational Psychology, Teachers' College, Columbia University

The task of education is to make changes in human beings. We teachers and learners will spend our time this year to make ourselves and others different, thinking and feeling and acting in new and better ways. These classrooms, laboratories, and libraries are tools to help us change human nature for the better in respect to knowledge and taste and power.

For mastery in this task, we need definite and exact knowledge of what changes are made and what ought to be made. In proportion as it becomes definite and exact, this knowledge of educational products and educational purposes must become quantitative, taking the form of measurements. Education is one form of human engineering and will profit by measurements of human nature and achievement as mechanical and electrical engineering have profited by using the foot-pound, calorie, volt, and ampere.

Until very recently, measurements of human qualities in education were rare. For example, the educational measurements reported by the federal and state and municipal governments up to 1910 concerned chiefly time and money, the number of teachers and students engaged, the number of days they spent, the value of buildings and grounds, the cost of books and supplies. The abilities of those who were educated and the betterments of intellect, character, and skill which were produced in them were left to speculation and faith.

We had, of course, alleged measures of educational achievement in the "marks" or "grades" reported for each student in each study or activity, in promotions and graduations and honors, and in the results of examinations for licenses to practice law and medicine, or to teach, and for various posts in the civil service. These marks and grades, however, were opinions rather than meas-

urements, and were subject to two notable defects. Nobody could be sure what was measured, or how closely the measure tallied with the reality! Marks in freshman algebra, for example, might be measures of inborn talent for mathematics, or of acquired power at mathematics, or of mathematical erudition, or of temporary memory, or of docility and fidelity in doing what the instructor ordered, or of sagacious divination of what the instructor desired! When we measured length or weight or volume or temperature or electric potential, all competent persons measured the same thing. But when we measured achievement in first-year Latin or college algebra, even the most competent twenty teachers measured twenty different composites.

Dearborn found, for example, among instructors teaching the same subject in the same college to the same grade of students, some who gave ten times as many "A's" as others did, and reported less than one-tenth as many failures. Finkelstein found that identical students in the same course taught during the first semester by one instructor and during the second by another, had three times the probability of a mark above 85 in the one case that they had in the other.

The general result was scandalous. Foster found in the elementary courses at Harvard that "A's" were thirty-five times as common in Greek as in English. Meyer found that over a period of five years one professor had never permitted a single student out of nearly a thousand to fail, whereas another in the same college reported nearly three hundred per thousand as failures.

Moreover, even when we did know fairly well what we were measuring, the mark or grade given by any one examiner might correspond only by a shockingly wide margin with the reality. For example, let the ability to be measured in geometry be defined as the ability to answer a certain specified set of questions and prove certain specified propositions. Elliott and Starch found that a hundred experienced teachers of mathematics assigned grades ranging from 28 to over 90 to the same set of replies in an actual examination paper.

It may be thought that such variations as this 28 to 90 are largely due to a general severity or leniency in the judge, in which

case deans, scholarship committees, and even students, might allow for them by multiplying each instructor's marks by some quantity representing his personal equation. The more important factors in causing such variations are, however, variations in the importance assigned to different qualities and a sheer inability to judge educational products accurately. Allowance for personal severity or leniency fails to eliminate the variation or greatly to reduce it.

When a student received 70 as the official rating of his work for a year in English composition or Elementary Chemistry, or the History of England, neither he nor we knew what it was 70 of, nor whether it was really 60, 65, 70, 75, or 80 of it. Clearly defined units of measure and instruments by which to count them were lacking.

The first steps to establish such units of educational products, and to devise instruments to measure them with reasonable precision were taken about a dozen years ago. The work began naturally enough with the simple matters of reading, writing, spelling, and arithmetic, which are a large fraction of the task of fifteen million children in this country each year.

The hypotheses and experiments involved in establishing such educational units and scales are somewhat intricate and elaborate, and are too technical for presentation here, but the nature of the scales themselves may be at least roughly illustrated.

In penmanship, for example, imagine a row of specimens of handwriting beginning with one called zero because it is just not legible and possesses just not any beauty or other merit in handwriting. At the other end of the row is a specimen called 17 which possesses a very large amount of general merit as handwriting. In between are specimens representing 1, 2, 3, 4, 5, and so on, each step of difference in merit being equal to any other. The unit is one-tenth of the difference between the best and worst writing found in 1000 children of grades 5 to 8.

When a desired or obtained change in ability to write is defined as improvement from 8 to 10 in this scale, anybody, anywhere at any time, can know what is meant almost or quite as definitely as when we speak of a baby changing from 8 to 10 pounds in weight, or a current increasing from 8 to 10 amperes. Impartial

judges, rating a pupil's handwriting by pushing it along the scale until the point is found which it most resembles, will agree closely—not, of course, as closely as they would in measuring a wire with a foot-rule, but, with the aid of repeated measurements of it, closely enough for any important educational purpose involved.

Or consider a measurement of word knowledge like this. The student sees a word followed by five other words or phrases. He is to underline that one of the five whose meaning is the same, or most nearly the same, as that of the given word. The test begins with words in the first thousand for importance, such as:

afraid	full of fear	possible	necessary	raid	ill
baby	manner	trembling	little child	notice	soft

It continues with words of less and less importance, but all in the first ten thousand for importance, having, for example, to represent the tenth thousand, such words as:

ambiguous	offensive	uncertain	roomy	very large	material
canyon	menagerie	palate	valley	gun	rule
classify	arrange	pacify	make clear	recede	promote
divulge	different	common	tell	repress	project

Such an instrument for the measurement of word knowledge has many merits. For our present purpose we may note two obvious ones: the score is absolutely objective—the same test paper would receive the same rating from any examiner; the examinations for different classes or in different years can be made exactly equal in difficulty.

While scientific workers in education have been establishing units and scales of educational achievement, the psychologists have been improving their tests of intelligence. The two sciences are also cooperating in devising tests of various scholarly capacities, such as the capacity to learn arithmetic, the capacity to learn to spell, or the capacity to learn Latin.

Measurements of pupils' capacities and achievements in more or less standardized psychological and educational units, are now a common feature of elementary schools. At least a million boys and girls, probably, were measured last year in respect to general intellectual capacity for school work. The number of such measures of reading, writing, spelling, arithmetic, history, and geography made during the year, probably exceeded two millions.

When we have measured a pupil in respect to his achievement in a school subject, and his capacity for that subject, the quotient of achievement divided by capacity is an important measure of accomplishment. A score of 70 made by a capacity of 70 is obviously very different from a score of 70 made by a capacity of 140.

In elementary schools, which are managed scientifically, these accomplishment quotients or ratios, familiarly known as A. Q.'s, are recorded year by year for each pupil. The pupils of great natural ability are required to do enough more than the average to keep their A. Q.'s near 1. They are thus protected against habits of idleness and conceit. The pupils of little natural ability are not rebuked or scorned for failures in gross achievement. They, too, are required simply to maintain their A. Q.'s near 1.

It may be expected that measurements of achievements and capacity and their quotients will soon be developed for use in high schools, colleges, and professional schools. It surely is unwise to have the measure of college students' achievement in English composition, or trigonometry, or beginning chemistry, or economics or second-year French depend upon the caprices of a thousand different individual instructors, if by enough ingenuity and care we can devise tests that will measure their achievements uniformly and precisely. The present condition at its best is shocking. The average correlation between the grades given in a subject and a student's real achievement in it is, in even the best American colleges, almost certainly not over .80, which means that the official ratings are six-tenths as erroneous as would be the case if the grades were assigned at random by a child, as in a lottery! If 900 students pass and 100 fail by the official ratings in a subject, there is every reason to believe that nearly half of those who failed really did better than some of those who passed.

It is demoralizing to students to find that their official ratings (on which degrees, honors, and financial rewards are given) depend so little on real achievement, so much on irrelevant matters and mere chance. It may, of course, be explained to them that, although any one mark is largely composed of error, the average of the score of marks received in two years will be a just measure of achievement in general. But such a lesson in the theory of proba-

bility gives little comfort to the student who has failed in subject A and must repeat it, though he had a much better mastery of it than of subject B in which he passed, or than another student had who passed in it.

As for the instructors, I do not know which is worse, the stupid conceit which assumes that the "A's" and "B's" and the "C's"—the 60's and 70's and 80's—are infallible indices of achievement and merit, or the sardonic indifference which prepares examinations whose findings it does not trust, and rates them carelessly with the excuse that even with care the ratings would be of little value.

That standardized examinations and other instruments for measuring achievement in colleges and professional schools are both possible and useful seems certain from experimentation of the last few years, slight as it is.

Their preparation, however, requires the cooperation of experts in the teaching of each subject and experts in mental measurement, a high degree of inventiveness, and much experimentation. Measuring achievement in a course in chemistry is a more elaborate task than measuring the atomic weight of oxygen. To measure improvement in knowledge of economics is harder than to measure the changes in the value of the dollar. Adequate units and scales for ability to read Latin may be more complex than Latin syntax itself. It may be many years before we can really measure achievement in, say, first-year French, so as to list its various features, define 0, 1, 2, 3, 4, etc., of each feature, know that what we call 4 of it is twice what we call 2 of it, and be able to tell with surety what amount of each any given student had at the beginning of the course and at its end. Until we can do so, however, all reports and grades are cryptic and likely to mislead; all comparisons of institutions and methods of teaching are insecure; all exact knowledge of what educational effort produces, is lacking. So it is our duty to try.

Moreover, every step of progress toward a truly objective measure is profitable. Last year, for example, those instructors in Columbia University concerned with the required freshman course in Contemporary Civilization, with some aid from an expert in mental measurement, prepared an instrument for testing achieve-

ment in that course, which took one step toward a genuine measure in place of opinion. It seems certain that none of the instructors and few or none of the competent students would be willing to go back to the old form of examination.

The case is nearly or quite as strong in measures of capacity. It surely is unwise to give instruction to students in disregard of their capacities to profit by it, if by enough ingenuity and experimentation, we can secure tests which measure their capacities beforehand.

Measures of special capacities, as for mathematics or for languages, have not, to my knowledge, been used as yet above the high school. But measures of general abstract intelligence or scholarly capacity have within three years come into wide use in universities. At about the same time, the Dean of Columbia College, the Director of Admissions in this University and Professor Colvin, of Brown University, began to take a careful measurement of general capacity to handle facts and symbols as one feature of the record of entering students.[1]

This measurement has abundantly proved its worth. It gives a very close prophecy of the grades a pupil will obtain in his freshman year—six-sevenths as close as one-half of the grades prophesies the other half. It points out almost unerringly any very stupid boys who have been hauled into college by their teachers' skill and their parents' money; or who have floated into college by careless certification. It helps the faculty or dean to decide quickly and correctly whether a case of deficient achievement is due to physical, intellectual, or moral causes. It permits the computation and use of an approximate A. Q., or accomplishment quotient.

At a certain university, for example, all the students of high scores in the capacity examination are called into conference by the dean and it is made clear to them that anything below A and B is essentially a failure for them, as anything below D is a failure for their less gifted fellows.

---

[1]Short tests, to serve somewhat the same purpose, but less precisely, had been used elsewhere, notably at the Carnegie Institute of Technology; and voluntary tests of certain psychological capacities had been made by the department of psychology at Columbia as early as 1894 for any freshman applying.

Of measurements in professional schools, I regret that time does not permit me to do more than mention the very active and important movement to that effect in schools of engineering stimulated by the Carnegie inquiry and its report of three years ago.

On the whole, it appears that the effort to replace opinion by measurement in our ratings of the achievement of higher education will increase and spread rapidly. Indeed, it may soon need protection from over-extravagant hopes more than from hostile criticism.

In the elementary schools we now have many inadequate and even fantastic procedures parading behind the banner of educational science. Alleged measurements are reported and used which measure the fact in question about as well as the noise of the thunder measures the voltage of the lightning. To nobody are such more detestable than to the scientific worker with educational measurements.

There are three criticisms in particular which even sound and accurate measurement in university education must meet:

First, it will be said that learning should be for learning's sake, that too much attention is given already in this country to marks, prizes, degrees, and the like, that students work too much for marks rather than for real achievement. Whatever force this argument has, is towards abandoning our official measures of achievement or towards making them measures of real achievement. Students will work for marks and degrees if we have them. We can have none, or we can have such as are worth working for. Either alternative is reasonable, but the second seems preferable.

Second, it will be said that the energy of teachers should be devoted to making achievements great rather than to measuring how great they are. It is true that for many teachers and many students, it is wise to teach and learn as well as may be, leaving the results to faith and hope, or even charity. Moreover, there are gifted personalities to whom scientific and business-like procedures are alien and even odious, and who should not be required to measure what they are doing or even, in the ordinary sense of the word, to know what they are doing. Their genius is better than efficiency. There are, however, not enough of these to be more than a negligible

factor in, say, the teaching of freshman English or first-year anatomy or the Law of Contracts. Most of us need to know what we are trying to teach or learn, and how far we have taught it or learned it; most of us will be aided, not hindered, by instruments for measuring educational purposes and products.

Third, it will be said that only the baser parts of education can be counted and weighed, that the finer consequences for the spirit of man will be lost in proportion as we try to measure them, and that the university will become a scholarship factory, turning out lawyers and doctors guaranteed to give satisfaction, but devoid of culture. This is a part of the general fear that science and measurement, if applied to human affairs—the family, the state, education, and religion—will deface the beauty of life, and corrode its nobility into a sordid materialism. I have no time to present evidence, but I beg you to believe that the fear is groundless, based on a radically false psychology. Whatever exists, exists in some amount. To measure it, is simply to know its varying amounts. Man sees no less beauty in flowers now than before the day of quantitative botany. It does not reduce courage or endurance to measure them and trace their relations to the autonomic system, the flow of adrenal glands, and the production of sugar in the blood. If any virtue is worth seeking, we shall seek it more eagerly the more we know and measure it. It does not dignify man to make a mystery of him. Of science and measurement in education as elsewhere, we may safely accept the direct and practical benefits with no risk to idealism.

CHAPTER II

PRINCIPLES UNDERLYING THE CONSTRUCTION AND USE OF INTELLIGENCE TESTS

STEPHEN S. COLVIN
Professor of Educational Psychology, Brown University, Providence, R. I.

The rapid development and extensive use of so-called intelligence tests during the past few years is one of the most striking and interesting facts in the field of educational psychology and one of the most significant in the province of school administration. Not only are psychologists today giving a large measure of their attention to devising, improving, and applying mental tests, but teachers and school administrators are employing these tests more and more to determine the ability of school children to do school work. Indeed, there is danger at present that the movement in the direction of intelligence testing may grow out of all bounds; that it may be misunderstood in theory and erroneously and even harmfully applied in practice. It is with the purpose of making somewhat clearer the nature of intelligence tests and of pointing out their value and their limitations that this chapter is composed.

I. WHAT IS GENERAL INTELLIGENCE?

1. General Intelligence a Native Endowment

Intelligence testing is concerned in determining what psychologists have termed "general intelligence." Just what general intelligence is may easily be misunderstood, although there is a fair, though by no means a perfect agreement as regard to the significance of the term. By the word *general* is commonly understood an innate ability or group of abilities that lie at the basis of the acquired intelligence of an individual. *Intelligence itself is not inborn, only the capacity to become intelligent.* For this reason some writers prefer the term "mental tests" or "mentality tests" to the term "intelligence tests," since these writers mean

by mentality the inborn capacity of the individual to become intelligent, provided he has the proper environment in which his mentality can develop into genuine intelligence. General intelligence, or mentality, then is to be understood as a native endowment which makes it possible for the individual to become more or less intelligent on the basis of this endowment. If a child is 'born long' in general intelligence, then he may, under proper conditions, achieve high intelligence in his knowledge of, and contact with, the world and his fellows; if he is 'born short' in general intelligence, then, no matter how fortunate his surroundings, he will be doomed to acquire in contact with his environment only a modicum of knowledge and skill.

2. General Intelligence Either a Single Capacity or a Group of Related Capacities

While all competent authorities would agree that the expression "general intelligence" designates inborn capacity to acquire intelligence in the various situations of life, they would disagree as to the further interpretation of this term, in regard to the significance not only of "general" but also of "intelligence." There are some who hold that the word "general" signifies a single inborn capacity to become intelligent in all situations; others that the term "general" means nothing more than that a person is born with a large number of specific capacities, more or less related, which enable him to acquire intelligent behavior in many different activities. The supporters of this first view, notably Spearman, Hart, and Burt, explain innate intelligence as a "general common factor." Similarly, Pyle has attempted to show that all individuals have a certain all-round learning capacity which is constant for different types of material. He believes that children and adults differ widely in innate learning ability, irrespective of the material learned, and that this ability is identical with, or closely related to, general intelligence. The writers who urge that general intelligence is an innate central capacity think of it as a single quality that may be transmitted, as the color of eyes is transmitted, from parent to offspring. Individuals inherit this all-round unitary capacity, and if it manifests itself more in one kind of activity than

in another, this difference is not due to the fact that there are parts, or aspects, to general intelligence. The differences are due either to other inherited abilities or to the varying opportunities presented to the individual to learn in different fields of human activity. Specifically, if a child acts with great intelligence in his class in arithmetic and very stupidly in his class in music, this is not due to the fact that he had two kinds of innate intelligence, one for number and one for music, but rather to differences in opportunity to learn and interest in learning in these two fields, or to specific inborn capacities which in one instance favor the development of his general intelligence and in the other hinder this development. For example, no matter what the general intelligence of the child might be, he could hardly be expected to become highly intelligent in his work in music if he were born with a poor sense of rhythm and with an innate inability to distinguish between tones varying in pitch. In such a case his general intelligence would have little or no opportunity to manifest itself in the face of so specific an inborn handicap.

While there are some who strongly hold to the view above outlined—that general intelligence is a unitary or central inborn factor—there are others who take the view that the term designates a large number of more or less closely related innate capacities to become intelligent in various life activities. Thorndike, in particular, advocates this view. He holds to a multiplicity of innate abilities that are related in varying degrees. He believes that between desirable single traits in a single individual there is a positive relation. "Having a large measure of one good quality *increases* the probability that one will have more than the average of any other good quality." The fact that a child has pronounced native ability in arithmetic is an indication that he will have more than average native ability in geography, even that he will be above the average in his moral qualities, but it is not certain that he will be. According to Thorndike, then, general intelligence is a term by which a large number of innate abilities to become intelligent may be classified, or arranged in a pigeon hole for purposes of convenience, because all the abilities so arranged are likely to be in some kind of agreement. More specifically, Thorndike believes that there

are three main types of innate intelligence, namely, intelligence for words and abstract ideas; motor intelligence, or skill with the use of the hands, and social intelligence, or the ability to get on well with one's fellows. These three types are positively related, but not necessarily in a high degree. The first type concerns itself particularly with abilities necessary to get on in school and college in the ordinary academic courses and in the more abstract aspects of applied courses. The second type of ability concerns itself with the execution of skillful motor acts and the comprehension of mechanical constructions and processes. The third type has to do with the understanding of one's fellows and with influencing and leading them. In order to be an excellent mathematician or classical student one must be 'born long' in abstract intelligence; in order to handle tools deftly, to invent and design, one must have in a considerable degree the second type of intelligence; in order to be a successful salesman or a social leader one must possess superiority in the third type of intelligence.

Not only are there three main types of innate intelligences, but within these main types there are subdivisions. An intelligence test that surveys a person's general intelligence does not indicate in particular the various aspects of this intelligence. To quote Whipple[1]: "Take, for instance, the testing of the mentality of a gifted child, a Winifred Stoner or a William James Sidis. To discover by simply testing that such a child has an I. Q. of a given amount is interesting, but it fails to get us anywhere in our real inquiry as to just which ones of the various mental functions are possessed of the extraordinary heightened efficiency. Is it memory span or capacity for concentrated attention or ability to handle symbols or apprehension of abstract relations or acute perceptive capacity or lively imagination or originality or breadth of associative tendencies or speed of learning or what that demarcates such a child from other children? What about his special abilities: does his musical, mechanical, arithmetical, linguistic, dramatic, executive, poetic, artistic and so forth ability exhibit the same unusual development or not? These questions compel us to plan out an

[1]G. M. Whipple, *Bulletin of Extension Division, Indiana University*, "Fifth Conference on Educational Measurements."

elaborate program of mental testing and to carry this forward on the one individual until we can plot for him a comprehensive 'psychogram' or 'psychological profile.'"

Thus the question as to whether there is *a* general (innate) intelligence or various kinds of general intelligences, more or less closely related, in the same individual is still a matter of controversy. The writer, personally, is inclined to the second view. He is led to assume that there are various inborn abilities that are general in their character in the sense that they appear in many life situations and in a somewhat close agreement in a single individual and that at the same time there are abilities of a very specific character that are not closely related to other abilities. Generally speaking, a pupil who has the capacity to do good work in arithmetic or algebra is likely to stand well in history or geography or general science; he may do good work in the manual training shop, though this is by no means certain. It would not be safe to predict confidently in regard to his ability to sing or act, to paint or to dance, and it is quite possible that, while he might stand at the head of his class in high school or college, he would have little or no native ability as a newspaper reporter or a salesman. After all, to the practical schoolman it makes very little difference whether general intelligence is a central factor or a bundle of different abilities related positively; *the child cannot be treated as a unit—he must be discovered in his various tendencies and abilities* and if we wish to know him as he really is, we must be able to work out the "psychogram" which Professor Whipple has mentioned.

3. General Intelligence is Fundamentally, Ability to Learn

Up to this point our discussion has concerned itself with the significance of the term "general" as descriptive of intelligence. We have seen that it means an inborn capacity or group of capacities more or less closely related. All psychologists agree that it refers to something innate, something that cannot be acquired or learned. Some psychologists consider it to be a single, unitary, central trait, others a group of traits that can be conveniently classified together and which show certain relationships and correspondences. It is now left for us to consider what the second part

of the term "general intelligence" signifies to psychologists. Here again we find a reasonable, but not a complete, agreement.

Recently a group of fourteen psychologists, authorities on mental testing, contributed to a symposium on the subject of "Intelligence and Its Measurement" in the *Journal of Educational Psychology*.[2] In this symposium they gave their views as to the nature of general intelligence. Some took the ground that the term intelligence could not be adequately defined or described in the present state of our knowledge; others gave very broad definitions, such as the "power of good responses from the point of view of truth or fact," or "the ability of the individual to adapt himself adequately to relatively new situations in life." Some emphasized the rational element as the essential one, considering intelligence as the ability "to carry on abstract thinking." This latter definition doubtless concerns the highest level of intelligence, and is one very essential aspect of it, but an individual may have little ability to deal with abstract ideas or to reason and may still possess a modicum of intelligence. Indeed, the intelligence tests so far devised give only a small part of their attention to the testing of reasoning abilities, and devote a much larger share to more simple intellectual processes. Buckingham[3] seems to express the matter of intelligence tests and the nature of intelligence in a helpful way when he says that, whatever our views may be in regard to the nature of intelligence in the abstract, "we are justified, from an educational point of view, in regarding it as ability to learn, and as measured to the extent to which learning has taken place or may take place."

An inspection of the various intelligence tests now in use clearly shows that psychologists have accepted this definition practically, if not theoretically. Intelligence tests are by no means confined to problem-solving, even in its simplest forms. They determine an individual's intelligence largely in terms of what he has learned, thus obtaining a measure of his ability to continue learning. Vocabulary tests, range of information tests, same-and-opposites tests, tests of fundamental operations in arithmetic (one

---

[2]March, April, and May, 1921.

[3]*Journal of Educational Psychology*, Vol. XII, No. 5, p. 273.

of the most widely used) and the like, demand little that is novel, little that tests rational powers. If an individual has sufficient knowledge and skill he can pass these tests. They measure intelligence only on the assumption that they test ability to learn by discovering what has already been learned. Even those tests that involve ingenuity, deliberation, and choice with words or things are based on elements that show what a person has already acquired. An example of this fact may be shown by the following extract from a test:

> Below are five words, four of which are related according to some principle. One word is not so related. Cross out the unrelated word: *physics, chemistry, geology, ~~history~~, biology.*

Now it is quite obvious that a successful passing of such a test is in part dependent on an ability to reason, to classify, to meet intelligently a new situation, or on some other similar mental activity of a fair degree of complexity; but also a large, perhaps the greater part is dependent on a knowledge of words and their significance in more or less detail. This knowledge is based on previous learning. It is clear, then, that a considerable part of intelligence testing is dependent on what has been learned; further, it should be remembered that the ability to learn is very closely related to the capacity to meet new situations intelligently, to reason, to abstract, etc. Therefore, to identify general intelligence with native learning ability is, both theoretically and practically, justifiable. We shall not be far from the truth when we define general intelligence as *a group of innate capacities by virtue of which the individual is capable of learning in a greater or less degree in terms of the amount of these innate capacities with which he is endowed.*

II. How Can General Intelligence Be Measured?

General intelligence is an inborn capacity. It does not manifest itself, however, except through learning. If an individual were born with a very high capacity to become intelligent, but had no opportunity to learn, he would possess no intelligence. *Intelligence must be acquired. Only the capacity is inborn.* There has been much argument in recent years as to whether nature (inherited

capacity) or nurture (training of the environment) is the more important. The whole discussion is likely to be beside the point and quite misleading unless care is taken to define exactly the position taken by those who debate the question. It is quite evident that a feeble-minded child can never become highly intelligent, never mind how favorable his environment, how skilled and patient his teachers. His innate endowment will not permit him to go beyond a certain level of attainment. Water will not rise above its level. On the other hand, the greatest potential intelligence will never become highly intelligent in an environment that affords scant opportunity to learn. The brightest European child reared from birth by a group of African Pigmies would appear as a moron or worse if later transported to a highly civilized and cultured environment. Whatever the native mentality of a deaf-mute, that individual must actually grow up as feeble-minded unless special methods of instruction are employed to reach his native ability and develop it. The truth of the matter is that when an environment is practically the same for a group of individuals, then the great differences that are found among these individuals are due to differences in native ability. Specifically, if forty children in the fifth grade of the elementary school show varying degrees of attainment in their school work, it is probably true that these differences are to be explained to a considerable extent as arising from inborn differences in mental capacities. The justification for the truth of this explanation lies in the fact that all of these children have had similar opportunities and similar incentives to learn. The environment in which they have been reared, while not identical for all, has not varied substantially from child to child; at any rate they have had about the same schooling. One factor (the environment) in the acquisition of intelligence has been practically constant; hence differences in acquired intelligence must be largely due to the other factor (innate capacity to learn). *Nature is more important than nurture in explaining individual differences in acquired intelligence, when the nurture has been similar for the group concerned. On the other hand, it would be equally true that nurture would be more important than nature in explaining individual differences if the native equipment of a group were substantially the same and the environment markedly different.*

1. Mental Tests are Possible When Based on Elements Involving the Common Experiences of Those Tested

The foregoing consideration explains the feasibility of devising tests to measure general intelligence. At first thought, it may seem impossible to determine the amount and nature of an innate capacity or group of capacities that manifest themselves only through learning. These capacities can be measured only indirectly through what has been acquired, never in their native purity. However, they can be indirectly measured successfully by measuring the acquired capacities in a group with substantially the same experience. *We never measure inborn intelligence; we always measure acquired intelligence, but we infer from differences in acquired intelligence, differences in native endowment when we compare individuals in a group who have had common experiences and note the differences in the attainment of these individuals.*

2. The Binet and Subsequent Tests Constituted on This Principle

Hence it follows that an intelligence test, to be valid, must be composed of elements appealing to the common interest and within the common experiences of the group tested. All successful intelligence tests have implicitly or clearly recognized this principle in their construction. As a case in point let us consider the Binet tests as originally devised by their author. They show on examination the fact that their separate tests were arranged on the basis of the common experiences of the children of varying ages. Children failing to pass tests for their particular age satisfactorily were classed as subnormal because they were below the reasonable attainment of their group. In no case were tests employed that were based on peculiar conditions or unusual opportunities for learning. Tests for any given age are given on the assumption that all normal children should have learned the things with which they have had common acquaintance. For example, a child of three is asked to point to his eyes, his nose, his mouth, to tell what he sees in a simple picture, etc.; a child of four to identify a key, a penny, and a knife. An older child is asked to count and make change, to give a rough definition of certain simple objects, to execute brief commands, to estimate weights, to give explanations and reasons, to

make aesthetic comparisons, and so on. The validity of this mental examination is definitely dependent on the extent to which the children examined have had previous knowledge of the items in which they are tested. Clearly, a child of three, however bright, could not point to his nose unless he had previously learned about this part of his face. To count pennies, to make change, to give sensible answers and explanations, these attainments are conditioned on the opportunities the children have had to learn about pennies, actual practice in counting and making change, knowledge of the words which they are to define, etc. Binet found, for example, that the average child of seven years could do certain things and answer certain questions. If a child of seven falls far below the average in his ability to respond to the tests, this is not because of lack opportunities to learn, but because of definite inability to learn. Such a child is feeble-minded if this inability is pronounced.

3. Not Only is a Valid Mental Test Based on Common Experiences; It Must Assume Common Interests as Well

It cannot be too strongly emphasized that *no test to determine intelligence is valid unless the individual tested has had a reasonable opportunity to learn about the various elements involved in the test and has also been interested in learning.* Some errors have already been made and still more are likely to be made in drawing conclusions as to the absolute or relative intelligence of individuals in a group or in various groups when the experiences and interests of members of the group or groups have been to any considerable extent different. A few specific instances will make this important point clear. It is a striking fact that the Army Alpha Tests, which in the past few years have been given extensively in colleges, normal schools and high schools, show in practically every instance higher average scores for men and boys than they do for women and girls. The conclusion might be reached that the intelligence of men on the whole is somewhat superior to that of women. That such a conclusion is not justified is at once seen when the Alpha Tests are examined. These tests were devised to measure the intelligence of soldiers. They included materials which on the whole would be somewhat more familiar to men than to women, because

the *interests* of the sexes are not by any means the same. It is the *interest* here in learning rather than the actual *opportunity* to learn that determines whether the test is equally fair for both sexes.

Another and more emphatic instance in point will show even more clearly how the matter of interest may determine whether materials included in a mental test are equally fair for all tested. A few years ago the writer gave the Stenquist mechanical ingenuity tests to two high-school groups, one of boys and the other of girls. The boys scored decidedly higher than did the girls. The difference was impressive, and from it might have been concluded that the innate mechanical intelligence of the boys was vastly superior to that of the girls. The facts, however, warrant no such conclusion. Girls traditionally are not interested in things mechanical, and not being interested in them, they do not learn about them. They may or may not have equal innate mechanical intelligence. The Stenquist tests could throw no light on this problem unless they were given to groups of boys and girls all of whom had had the same opportunities and incentives to learn about mechanical facts and principles.

4. Scores Obtained in Typical Intelligence Tests Conditioned in Part on Knowledge of English

As has been said, opportunity to learn as well as interest in learning is a determining factor in devising and using mental tests. As an illustration of this may be sighted results obtained in giving the Otis Intelligence Tests to the children of the public schools in Brookline, Massachusetts, and in Cincinnati, Ohio. In the former city the tests were given under the direction of the writer; in the latter, by Warren W. Coxe. In Brookline the average scores were much larger than in Cincinnati. The children of Brookline were on the whole a clearly superior group, according to the published Otis norms, while the children of Cincinnati were somewhat inferior. An average Brookline child of twelve would have, according to the results of these tests, a mental age about two years in advance of the average Cincinnati child. Are we to conclude, then, that the Cincinnati children are really inferior in innate intelligence to the Brookline children? I am inclined to think not.

The great differences in the scores I attribute to differences in opportunities to learn words and their meanings. Examination of the Otis tests, and other similar tests, will show that success in passing these tests is conditioned largely on extent and accuracy of vocabulary and on verbal ingenuity. In no single element entering into school attainment do children vary so much as in the knowledge of words and the ability to use words. Much of this knowledge and skill is determined by the home environment. Brookline is, on the whole, a center of culture where the children acquire at home an ability to use English in a superior degree. The same is not so conspicuously true in Cincinnati.

That this explanation is not altogether fanciful is shown by the following facts: In Brookline there was a considerable difference in the median scores, as well as the maximum scores, for the children of the 'better' and the 'poorer' localities. These differences were marked in the case of most of the verbal tests; they were not found to exist when the arithmetic tests were examined. Clearly, the differences were differences in verbal ability, not in innate intelligence.

Further corroborative evidence that this explanation is at least in part correct is indicated by the circumstance that a number of students in Brown University either foreign born or of foreign extraction have received low scores on their mental tests but have done good college work. On investigating these individual cases, I have found that the low psychological scores are to be explained by the fact that these students have not the same familiarity and facility with the English language as those who have been reared in a more favorable environment. It is not their innate intelligence that is inferior, but their mastery of the vernacular.

Carrying this investigation somewhat further, I have collected data to show that in the City of Providence the Italian children receive scores in the National Intelligence Tests (largely verbal) on the average lower than those of the children reared in an English speaking environment. The Italian children, therefore, appear to be as a class of less intelligence than the children of native parentage. A more careful examination of these different groups reveals the fact that the National Intelligence Tests tend to under-

rate the real mentality of the Italian children. They score lower than the English groups because of a less familiarity with English. *It seems probable that all mental tests that are largely linguistic will be unfair to those persons whose training in English either at home or in the schools has been inferior.* It is only when individuals tested have had common opportunities to learn the vernacular that real differences in intelligence can be surely inferred from the scores secured. It must be kept in mind that *no general tests for general intelligence have yet been devised. Tests are valid only within a group who have had identical or very similar opportunities for gaining familiarity with the materials of the test, and who have not only the same opportunity to learn, but the same desire to learn.*

5. In Order to Secure Valid Results the Administration and Scoring of Tests Must be Uniform

Further, the validity of tests is based not only on the considerations pointed out above. It is likewise dependent on the care, accuracy, and consistency of administering and scoring. Tests poorly and carelessly given and scored may give one result; tests carefully and accurately given and scored quite another. Indeed, Coxe in attempting to explain the great differences between the Brookline and the Cincinnati scores says: "The only possible explanation that occurs to us is in the method of giving and of scoring. He then goes on to point out that the tests in Cincinnati were given with the greatest care by himself and one assistant. However, this explanation does not seem to account for the differences in this particular instance, since the Brookline tests were administered only after very careful instruction of the teachers in the method of giving the tests, and since the results showed consistency among themselves. If they had been given carelessly and in various ways, there would have been no general tendency in one specific direction, as was the case with the Brookline scores.

However, that the significance of tests may be greatly impaired by lack of uniformity and care in administering and scoring seems to be shown by the results that Book[4] obtained from a

[4]W. F. Book, *Preliminary Report of State-Wide Mental Survey of High-School Seniors,* Univ. of Indiana, 1920.

mental test given to the seniors in the high schools of Indiana. He sent to the various high-school principals of the state copies of the Indiana University Intelligence Scale, Schedule D (the Pressey Tests), through the offices of the state high-school inspector. With the test blanks were sent manuals of instruction to teachers and explicit directions for giving the tests. The actual giving of the tests was intrusted to a large number of individuals, many of whom had little or no knowledge of mental testing and few, if any, of whom had had any definite training in giving the tests. Under such conditions there must have been considerable variation in the manner in which the tests were administered. The result showed a low positive correlation between the scores in the mental tests and the previous school records of the seniors tested, as well as other facts that indicated that the relation between intelligence and school success was not so pronounced as is probably the case. Had these tests been more carefully and uniformly administered, it is certain that the findings would have been more definite and of greater practical value.

6. Summary

It may be seen from the foregoing discussion that in giving mental tests the following considerations should be definitely kept in mind:

1. Are the tests so devised as to be suited to the group tested? Particularly, do they contain materials with which all tested have had similar incentives and opportunities to gain familiarity?

2. Can comparisons safely be made between the group tested and other groups that have already been tested or are later to be tested? In other words, can general norms be relied on, or is it necessary to establish a norm for the particular groups tested? The writer's opinion is that *in the case of the great majority of the mental tests now on the market, little of definite value can be obtained by the use of the general norms already published.*

3. Are the tests administered and scored in a careful and uniform manner? Tests are much more satisfactorily administered if given by one individual trained for the work. When the tests are administered by a number of individuals there should be ample

discussion of the nature and significance of the tests and practice in their use before they are given.

III. Origin and Development of Mental Testing

1. Study of Individual Differences

The first extensive and practical test to measure mentality dates back to the pioneer work of the French psychologist, Binet, who collaborated with the French physician, Simon, in the first decade of the present century. Binet quite appropriately is considered the founder of the movement. However, in a very real sense attempts had been made to determine innate abilities several decades before Binet published his original intelligence scale. Individual testing arose with the study of individual differences, and is contemporaneous with the work of Sir Francis Galton. Galton's work in the direction of mental testing was largely made known and developed in America by James McK. Cattell, as Professor of Psychology in the University of Pennsylvania and later in Columbia University. Cattell's service in the field to mental testing is well stated by his most distinguished pupil, Professor E. L. Thorndike. Of this work Thorndike says:[5] "Cattell refined Galton's methods and won recognition for such measurement of individuals as a standard division of psychology and of psychological training in universities, beginning at Pennsylvania the systematic inventory of mental traits which became such an important feature of the Columbia laboratory and which was for so many of us an introduction to the whole topic of individual psychology. His paper of 1890 on 'Mental Tests and Measurements' (*Mind,* Vol. 15, pp. 373-380) was the first of a series of influential contributions made during the decade and associated primarily with the names of Kraepelin, Binet, Cattell and Jastrow." On referring to this early paper of Cattell, we find a description of the tests used by him and the statement that some of these had already been used by Galton in his Anthropometric Laboratory at South Kensington Museum. An examination of Cattell's tests shows that they concern

[5]*Columbia University Contributions to Philosophy and Psychology,* Vol. XXII, No. 4 (1914); p. 92.

themselves largely with sensory discrimination and rapidity of reaction. Likewise immediate memory (memory span) is tested by finding the number of letters a subject remembers at one hearing. Ability to estimate space is determined by a test requiring the bisection of a line of 50 cm.; ability to estimate time is tested by estimating a ten second interval. A judgment of least noticeable differences in weight is also included. In a later article by Cattell and Farrand[6] we find a description of the further extension of the work of mental testing as employed with students of Columbia University as subjects. The tests used included handwriting, visual acuity and color vision, auditory acuity and perception of pitch, sensitivity of the skin, perception of weight, sensitivity to pain, accuracy and steadiness of movement, reaction time, cancellation of A's, perception of time and space, memory-span, memory of length of a line previously drawn, after-images and mental imagery. In regard to these tests Cattell says: "Our experience with these tests leads us to recommend that they be made a part of the work of every psychological laboratory."

It can be seen that these earlier attempts at mental testing concerned themselves chiefly with what may be designated as the sensory and motor phases of mentality, and gave scant notice to the more elaborate phases of intelligence. In the tests of Binet we find several that are identical with, or similar to, these earlier tests. Specifically, we find in Binet's scale, memory-span test (in this case for digits and for words in a sentence rather than for letters); a test involving the estimation of space; another involving judgment in regard to weight. In addition to such tests as these the Binet scale includes tests regarding familiarity with common objects, tests that involve comparison and judgment on a rather high level and so on.

2. Binet's Scales and Their Revisions

Binet's first scale appeared in 1905; it included thirty tests and was roughly standardized. The scale of 1908 comprised fifty-six tests, arranged for the ages from three to thirteen. This scale was

[6] "Physical and Mental Measurements of the Students of Columbia University," *Psychological Review*, Vol. 3, pp. 618-648 (1896).

revised and republished in 1911. In this final revision by Binet there were five tests arranged for every year, except one, from three to ten. Tests for the ages of twelve and fifteen were also included. Goddard, then at Vineland, used Binet's scale in dealing with his subnormal children. He also measured 2000 normal children with these tests, publishing the results in the *Pedagogical Seminary* for 1911. The Binet tests have been extensively used in America for a decade, and in the course of this time they have been extended and revised. Goddard made some slight revisions, in his work at Vineland. In 1915 Yerkes and others published a point-scale revision of Binet's tests. Kuhlmann has also revised Binet's tests in his work with subnormal children at Faribault, Minnesota. The most extensive and fundamental revision has been undertaken and carried out by Terman. His results appeared in 1916.[7] A pupil of Terman, Otis, has also worked out a standardization of an absolute point scale on the basis of the Binet tests. Of the various revisions of the Binet tests, that by Terman is the most important. The "Stanford Revision" (as these tests are called) was "the result of several years of work, and involved the examination of approximately 2300 subjects, including 1700 normal children." There are ninety tests in all, six for each age level from three to ten, eight for the age of twelve and six for the age of fourteen. There are also six tests for average adults and six for superior adults. A number of alternate tests for the various ages were also provided. Of the thirty-six new tests twenty-seven were added by Terman; he also borrowed a few tests from other sources.

3. Methods Used to Designate a Child's Intelligence

Binet expresses the child's mentality by giving his mental age in relation to his chronological age. Yerkes in his point scale shows the same facts by giving the total points scored by the individual in comparison with the average points scored by normal children of the age of the child tested. For example, a child whose chronological age is ten, when tested by the common form of the Binet tests might show a mental age of eight. He would then be classified as two years retarded in mental age by Binet. In the Yerkes scale

[7] "*The Measurement of Intelligence,*" Boston, 1916.

the same fact would be expressed by the statement that he received a total score of thirty-nine (the average score for a child of eight years), while if he had been normal he should have received a score of fifty-nine (the average score for a child of ten years). His actual intelligence is indicated by the ratio of the score made to the average score of children of the same chronological age as the child tested.

Terman in his treatment uses a somewhat similar method of indicating the individual's mentality. He states intelligence in terms of the I. Q. (Intelligence Quotient), which is obtained by dividing the child's mental age by his chronological age. Thus the child above referred to, whose mental age is eight and whose chronological age is ten, would have an intelligence expressed by an I. Q. of .80. This method of indicating a child's mentality has certain points in its favor, but it likewise involves dangers which must definitely be guarded against when I. Q.'s are used for administrative purposes. The chief value of the I. Q. lies in the fact that it expresses the child's innate intelligence in a more or less absolute way. It is intended to indicate his actual mentality irrespective of his age. According to Terman, an I. Q. remains permanent (with possibly slight changes) throughout an individual's life, at least up to the period of old age, when mental impairment begins with the breaking down of bodily functions. This would mean that if a child of five chronologically was mentally four years old, he would have an I. Q. of .80; at ten years chronologically he should have a mental age of eight and still an I. Q. of .80. Terman's contention seems on the whole to be substantiated by the facts, although it is probable, in some instances at least, that a child's I. Q. may vary from year to year, and that at times it may have a tendency to increase and at times to diminish.

While the I. Q. serves a very useful purpose in indicating to the teacher and administrator the probable intelligence of the pupil at each successive stage of his school progress and is important in forecasting the character and extent of his school attainment, it should never be used for purposes of classification of pupils without also taking into consideration the actual mental and chronological age of these pupils. This, of course, is a matter of plain

common sense, but a word of caution may not be out of place, particularly since in certain instances pupils have been compared and classified in their school work on the basis of I. Q.'s alone. Yet it can clearly be seen that children of the same I. Q. may be far apart in actual school attainment, because of differences in mental and chronological ages. *Children of varying mental ages, and even children of similar mental ages, but of markedly varying chronological ages, cannot be safely grouped together for school instruction. Innate intelligence, considered by itself, does not give us information in regard to acquired intelligence. We must group children for instructional purposes largely on the basis of their acquired intelligence and to a lesser degree on the basis of their chronological age. However, children who are approximately of the same mental age and whose chronological ages are not markedly different may be safely classified according to their I. Q.'s.*

The Binet tests were worked out by their author for the express purpose of segregating for special instruction all of the mentally defective children in the schools of Paris. Their aim was to detect feeble-mindedness. This original use, though still of importance, is of very much less value than their use in dealing with children of normal and supernormal mentality.

Various criticisms have been brought against the Binet tests, one being that they fail to be of any great service in accurate diagnosis of feeble-mindedness. Dr. Fernald[8] writes: "The Binet tests corroborate where we do not need corroboration, and are not decisive where the differential diagnosis of the high-grade defective from the normal is in question." This criticism is doubtless valid to the extent that the Binet tests are not suitable instruments alone to determine small variations in degrees of feeble-mindedness. However, they are on the whole reliable for discovering among school children those who are markedly deficient in intelligence, and they should be used for this purpose as well as for the classification of normal pupils. The Binet tests have been criticised also because they are too verbal in their nature; because they rely too much on words and too little on activities, *i. e.*, they appeal too much to abstract intelligence.

---

[8]*American Journal of Insanity*, 1914.

4. The Performance Test

Another type of intelligence test has been developed which in part at least meets these two objections to the Binet tests. This is the performance test, which like the Binet test, was worked out first for the purpose of detecting and diagnosing feeble-mindedness. The "performance test" is not, as is the Binet test, the work of a single individual; neither does it designate a specific group of tests. It is rather the name of a type of test or a method of procedure in testing. As the name indicates, a performance test emphasizes doing in a rather objective sense, generally doing with the hands. The intelligence of the individual is determined by what he does in response to a direction or command. Such a test may of course be executed with pencil and paper, but in its inception it was distinctly of the hand type of execution, with no writing or marking on paper involved. A test of this type is not only valuable as a supplement of the more verbal type of test, but is absolutely essential in determining the mentality of non-English speaking children, children with a limited English vocabulary and children with speech defects.

A common type of performance test is the form-board. This test originated with Seguin, and was employed in his work with mental defectives. It has passed through various adaptations, but its essential character has not been materially changed. It consists in fitting wooden blocks of various shapes into forms cut out to receive them. The board may be very simple, or it may be made as complex as desired, not only as to the shape and number of forms used, but also in regard to the blocks to be fitted, since each block may be a single solid piece or composed of a number of pieces. in which case the pieces must themselves be fitted together as well as placed in the proper form. A variation of this test consists of a puzzle in which various parts of a figure or shape are required to be fitted together, as, for example, in the Healy manikin puzzle. Picture puzzle tests have been largely used in recent years as performance tests. In this type of test the various parts of a picture are to be arranged in their proper order. In some instances a picture with parts omitted is given the subject, and he is required to complete the picture by filling in the gaps with the

proper blocks. Another type of picture test consists in arranging a series of pictures in such an order that they tell a complete story. A form of the performance test that is now frequently used is the "maze test." This test was used extensively twenty years ago, in the earlier days of animal psychology when the intelligence of an animal such as a white rat was studied by finding how easily and surely the animal could learn to go through the passages of a maze and get to the center where the food was placed. The Porteus[9] Maze Test for detecting feeble-mindedness is the best adaptation of this test. The maze test when used with human beings is a paper and pencil test of the performance type. The maze is printed on a sheet of paper, and the person tested is required to trace with a pencil the correct way of going through the maze. The form-board test and the various picture puzzle tests have also been adapted to paper and pencil use, but nevertheless retain their essential characteristics as performance tests.[10] Reference has been made to the fact that the performance tests have been adapted to the pencil and paper type of test. One reason for this adaptation is that the test may better be done on pencil and paper than as an actual objective performance. This would be true of the maze test primarily. It is more advantageous on the whole for the subject tested to trace the passages of a maze than to go through an actually constructed maze. It requires a kind of planning and foresight not so easily brought into play in the actual maze. Further, it is much more economical and easily administered.

However, the main reason for reducing the performance test to the paper and pencil form lies in the fact that by this means it can be made a group test rather than an individual test. Now it is quite clear that group tests are necessary in determining the intelligence of large numbers of school children. Individual tests require an enormous amount of time in their actual administration. Further, the difficulty of giving individual tests is very much

---

[9]This test, together with that of the Binet-Simon Scale, can be conveniently found in a handbook by N. J. Melville, *Testing Juvenile Mentality,* Second Edition, J. B. Lippincott Co., Philadelphia.

[10]A convenient description of some of the most important performance tests, together with methods of administration and results secured, is found in a book by Rudolf Pintner and Donald G. Paterson, *A Scale of Performance Tests.* D. Appleton & Co., N. Y., 1917.

greater, since they require an elaborate technique and a large amount of training on the part of the one who administers them. Group tests can be administered much more easily, and although the person who employs them should never do so without thoroughly understanding their nature and purpose and without careful training in the exact methods of administration, still the preparation required may be measured in days rather than in months.

5. The Development of Group Tests

The development of group tests is of a very recent date. The group tests originally were composed of materials of the verbal type rather than of the performance type and they still continue to be predominatingly verbal, though by no means exclusively so. Necessarily, group tests with children in the primary grades must be of the performance type, and it is advantageous to include in the test of older children some of the performance type.

In the early days of mental testing there was no pronounced call for group tests, since the necessity of testing large numbers of children for the purpose of classification and instruction was hardly recognized. The need was first felt, not in the school, but in the army during the emergencies of the World War. Immediately after the declaration by the United States of hostilities against Germany the American Psychological Association appointed various committees to consider what the psychologists of the country could do to aid the Government. One of the services rendered was the devising of a number of psychological examinations that were later applied to nearly two million men in the American army. Two types of group tests were finally worked out, one known as the Alpha test and the other as the Beta. The Alpha test was verbal in its nature and was employed in testing literates; the Beta test was of the performance type and was designed for illiterates and those who were unfamiliar with the English language. In addition to the group tests nearly eighty-five thousand men were given individual examinations. These individual examinations were the Point Scale, the Stanford-Binet and a Performance Scale examination. The army tests soon proved their worth as an aid in classifying soldiers according to their abilities, in detecting

and segregating or rejecting men of low military value, in prognosticating success of candidates in officers' training camps and the like. Soon after the signing of the Armistice the Alpha tests were made public and in the year following the end of the War were used to test students in a large number of universities, colleges, normal, and high schools. The success of these tests resulted in the construction immediately of a number of group tests of the verbal type for use in schools and colleges and also a little later of group tests of the performance type for use in the primary grades of the elementary schools. The verbal tests have in many instances included one or more tests of the performance type.

6. Characteristics of Present Group Tests

Although the Army tests furnish the first instance of the careful preparation, standardization, and use of group intelligence tests, scattered attempts had been made prior to 1917 to employ such tests in an experimental way. The framers of these earlier group tests and of the Army tests were not without guidance in their work. There were, in the first place, suggestions from Binet and those who had revised his work, particularly Terman. Few of the tests in the original Binet scale or in those of later revisions have been taken over bodily into the group intelligence tests, with the exception of those group tests worked out by Terman and Otis, but the principles and the fundamental characteristics of many of the Binet tests have been employed in making group tests. For example, in the Alpha examination the first test is a directions test; an important test in the Binet scale is the determination of ability of the child to execute a series of commands. The second Alpha test is an arithmetical problem test; Binet's original test involved counting and making change, and in Terman's revision we find an arithmetical reasoning test. The third Alpha test consists in selecting from three possibilities the best reason for a statement; while the Binet examination contained no test of this exact character, it provided various simple tests to determine the child's reasoning abilities. The fourth Alpha test presents a list of words associated in pairs. The subject is to determine whether these words are associated by the principle of likeness or opposition. The

Binet examination contained a free association test in which the child is required to name all the words he can think of in three minutes. Test five in the Alpha series is a disarranged sentence test. Words are given out of their proper order and they are to be put in the order that will give them sense. This is almost identical with one of the original Binet tests. Test six of the Alpha examination is a number completion test in which a number series is to be filled out according to the principle indicated in the part of the series given. This has no direct counterpart in the Binet series, which, however, uses counting, both forward and backward, as a test for intelligence. Number seven of the Alpha group is an analogies, or mixed relations, test which has no clear counterpart in the Binet tests. Number eight of the Alpha group is a range of information test; a number of the Binet tests are of this general type, though not of the specific form used in the Alpha test. In the Beta group the test that most closely resembles a Binet test is the picture completion test—a test that requires the addition of parts lacking in the picture.

Although those who have compiled group tests have, then, received substantial aid from Binet and his followers they have obtained help from other sources, notably from the tests devised by psychologists for the purpose of measuring individual differences. Mention has already been made of the work of Galton in England and Cattell in America, whose investigations, as has been pointed out, were primarily along the lines of testing the motor and sensory phases of intelligence. On the whole, the most important intelligence test contributed by psychologists for determining individual differences is the Completion Test of Ebbinghaus, devised by its author in 1905 for the purpose of investigating the fatigue of a school day in the City of Breslau. The original test consisted of a paragraph in which words with syllables omitted were presented to the subject, who was required to fill in the omissions. Terman, in his work with Childs on a revision and extension of the Binet Scale, published in 1912[11] a modification of this test in which a mutilated paragraph was prepared with four progressive degrees of difficulty. In this paragraph whole words were omitted rather

[11]See *Journal of Educational Psychology*, Vol. III, p. 199.

than syllables. Terman says that this test appears "to bring to light fundamental differences in the thought processes." He found the principal objection to the test to be the difficulty of standardizing it. Such a standardization has since been worked out by M. R. Trabue in his Completion-Test Language Scales.[12] This scale has further been restandardized by T. L. Kelley. In its present form it seems to be one of the most reliable single measures for intelligence that we possess. It is particularly suitable for determining some of the more complex forms of mental ability.

Although Terman was instrumental in improving the completion test, he does not include it in the Stanford Revision. The nearest approach to this test is his dissected or disarranged sentence test. Of it he says, "This experiment can be regarded as a variation of the completion test. Binet tells us, in fact, that it was directly suggested by the experiment of Ebbinghaus. As will readily be observed, however, it differs to a certain extent from the Ebbinghaus completion test. Ebbinghaus omits parts of sentences.... In this test we give all the parts and require the subject to relate given fragments into a meaningful whole."

Another test suited for discovering some of the more complex forms of intelligence is the Analogies, or Mixed Relations, test first used a decade ago by Cyril Burt in England. This test consists essentially in presenting three words in a series, the first and second of which bear a certain relationship. The examinee's task is to supply a fourth word that bears the same relationship to the third word as the second does to the first. The test is usually stated in the form of a proportion, thus: *Admire: Friends:: Detest:———?* The analogies test is frequently adapted to the abilities of little children and illiterates by substituting pictures for words.

The analogies test is a sample of a large group of tests, classified under the general name of "association tests." Some of these tests in their origin date back many years. As early as 1899 we find an article by J. McK. Cattell and Sophie Bryant on "Mental Association Investigated by Experiment."[13] The uncontrolled as-

[12] *Teachers College Contributions to Education,* No. 77, 1916.
[13] See *Mind,* Vol. XIV, pp. 230-250.

sociation method was used by Binet in testing how many words a child could name in three minutes. Controlled association tests are frequently used to-day in group tests of a verbal character. They include, besides the analogies test, associations of part with whole or *vice versa* (example, *chair–leg*); the genus with the species, or the reverse (example, *man–Indian*); a word with its opposite (example, *love–hate*); and other more complicated relationships. One of the most important of such relationships now frequently employed in group psychological testing may be designated as a classification test of which the following is an example:

Think how the first three words below are alike and then underline the one word of the last five that most resembles the first three: *ivory, snow, milk—butter, rain, cold, cotton, water.*

This test can easily be varied by substituting pictures or designs for words.

The substitution test, which determines the rapidity and accuracy of learning by substituting for one set of characters another according to a key, is also found in group intelligence tests. The intelligence of the person is tested by determining the progress made in learning to make these substitutions. Dearborn,[14] in 1910, describes such a test in an article discussing experiments in learning. In Dearborn's experiment numbers were substituted for letters combined into words in one test, and in another symbols were substituted for numbers. Dearborn names this test a "practice experiment" and he plots curves of learning based on the scores obtained.

Vocabulary tests, which are sometimes employed in the group tests of to-day, have been used by psychologists for many years. As early as 1891 Kirkpatrick investigated the "number of words in an ordinary vocabulary."[15] In more recent years Kirkpatrick has extended his investigations, and important studies have been made by Whipple, Ayres, and Babbitt among others. Terman included a vocabulary test in his revision of the Binet Scale and finds that this test shows a fairly high correlation with intelligence. The

[14] *Journal of Educ. Psychology*, Vol. I, pp. 378-384.
[15] *Science*, XVII, pp. 107-8.

vocabulary test is in reality a form of the range of information test now frequently employed in group testing.

Psychologists have given a good deal of attention to various forms of memory testing, but these tests play an inconspicuous role in the group tests of to-day. Rote memory, in particular, does not seem to bear a very close relationship to the more significant aspects of intelligence, though, of course, memory is basal to all learning.

The directions test as a response to verbal commands was, as we have seen, used by Binet in his scale. As a paper and pencil test it was put into form sometime before the war by Woodworth and Wells.

The cancellation test, in which certain digits or letters of the alphabet arranged in irregular order on a page are crossed out, has engaged a considerable share of the attention of psychologists, but has exhibited practically no relation to intelligence in its more developed forms. It is not employed in group tests at present.

Although the great majority of the mental tests found in the group tests now in use have been derived more or less explicitly from the work of Binet and other psychologists, two frequently employed tests at least are directly connected with attainment in school subjects. One of the common group tests now used is an exercise in the fundamentals of arithmetic or in simple arithmetical problems. The test involves concentrated attention, mental alertness and a fair degree of rational ability in some instances. The scores obtained show a fair degree of relationship to general intelligence.

The reading tests, particularly as worked out by Thorndike,[16] measure successfully some of the higher mental abilities. This test is of course very definitely related to one of the most essential requirements in school progress, namely, the ability to grasp and analyze the meaning of the printed page.

---

[16]Thorndike tests reading ability by requiring the subject of the test to read a paragraph and then answer certain questions concerning it with the paragraph still before him. Other reading tests of this character involve the reproduction of a paragraph from memory after the reader has perused it for a definite length of time.

This brief description of the origin of some of the most important elements in the group tests now used gives a general idea of the level of intelligent response required in performing the tests. It will be seen that on the whole the more complex factors of inference, judgment, and logical analysis are not extensively involved. An examination of nine of the most commonly used group tests shows that the most frequent single test is that of range of information, involving no rational ability; another favorite test is that of fundamental operations in arithmetic or the solving of simple problems. The opposites test is likewise frequently found. Among the more difficult tests logical selection and classification are often employed, as well as sentence completion. The analogies test is used in three of the nine sets.

IV. Intelligence and Character—Character Tests

It has already been pointed out in this discussion that intelligence tests measure not only intellectual ability, but also opportunity to learn and interest in learning. There are several other factors involved in the ability to perform these tests. Chief of these is the "will-to-do," the capacity to hold the mind down to a task and keep the attention alert and concentrated in the face of outside interests and distractions. The will-to-do is, to an extent, involved in the execution of an intelligence test, particularly if it is at all difficult and extended in scope, since the willingness to hold the mind to a task is here concerned. But it is not only in the performance of the test that this factor enters. It plays an important part in the acquired ability which enables the person tested to comprehend the materials presented, for, as has already been said, an intelligence test to a considerable degree measures ability to learn by measuring what has already been learned, and this acquired knowledge has been gained not merely through intelligence but through willingness to work as well. A child's success in school is due to his intellectual endowment in part, but only in part. His character and temperament are likewise important factors in his success or failure. Will-to-do a task bulks large in the total school performance. So it would seem that the present so-called intelli-

gence tests are in a measure character tests as well, but of course only in a very small and limited degree.

1. The Will-Profile Test

The attempt to determine character as independent of intelligence is scarcely in its beginnings. However, two fairly extensive character tests have been so far devised. The first of these is the so-called "Will-Profile Experiment" of Professor June E. Downey, of the University of Wyoming.[17] It is described as a tentative scale for measurement of the volitional pattern. It is for the most part a study of the variations of the handwriting of an individual under diverse conditions. Among the factors said to be tested are: speed of decision; the coordination of impulses under the mental set of both speed and accuracy; freedom from inertia as shown in speed in warming up, ability to maintain a high speed, etc.; ability to inhibit a motor impulse; flexibility of movement as shown in ability to disguise and to imitate handwriting; care in details; amount of motor impulsion; assurance; resistance to opposition; and perseverance. It is quite evident that this list includes a number of general characteristics that show the nature of the will of an individual. Through a single motor expression (handwriting) appearing in an experimental situation, conclusions are drawn as to the will tendencies of the individual as a general factor. These tendencies are supposed to express themselves in concrete situations.[18]

2. The Voelker Test

In contrast to the general character of the experiments of Professor Downey is the very concrete investigation of Dr. Paul F. Voelker,[19] who attempted to find out the truthworthiness of boys in actual life situations. Among the qualities that he has sought particularly to measure are: tendency to exaggerate; suggestibility; willingness to receive help in the solution of a problem when

---

[17] *University of Wyoming Bulletin,* Vol. XV, No. 6A (1919).

[18] An adaptation of this test has been worked out by the Bureau of Personnel Research, Carnegie Institute of Technology and published as Test IX.

[19] See *Religious Education,* Vol. XVI, No. 2 (1921), pp. 81-83.

such help is forbidden; punctuality in returning a borrowed object according to a promise; honesty in money matters as indicated by whether the boy will keep over-change given him in purchasing an article; willingness to accept a "tip;" his truthfulness under various conditions, and so on. Dr. Voelker found that the scores obtained by boys in these tests were largely influenced by instruction and environment. He found little agreement between a boy's intelligence and his standing in the tests for trustworthiness.

3. The Liao Tests

As another example of an attempt to determine character through specific tests may be mentioned the work carried on by S. C. Liao at Brown University. Liao prepared a moral judgment scale in the form of a "best reasons" test. A statement is made and under it are placed five reasons for the truth of the statement. The subject tested is required to indicate for every statement the best reasons. Under each statement one reason is moral in its nature, the other reasons being of a general or personal character. An example of this scale follows:

I. It is wrong not to work.
   1. Idle people are called lazy.
   2. Idle people earn no money.
   3. Idle people are discontented.
   X4. Idle people live on the works of others.
   5. Good men tell us we should work.

II. A kind word is better than a harsh word.
   X1. A harsh word makes others unhappy.
   2. A harsh word makes us disliked.
   3. President Roosevelt said, "Speak softly."
   4. A harsh word is generally a hasty word.
   5. Kind people succeed in life.

III. We should all try to get a good education.
   1. Educated people make the best citizens.
   2. They do better in business.
   3. They get the most out of life.
   4. Pupils are required to go to school.
   5. It is a pleasure to know a great deal.

IV. Our school is a fine school.
   1. The principal says it is.
   2. The teachers do not find fault with us.
   3. We are taught to help one another.
   4. We have a fine ball team.
   5. We are seldom punished.

V. If you have money you should give some to charity.
   1. It will make you feel happy.
   2. It will help those who are in want.
   3. Those you help will like you.
   4. People will think well of you.
   5. The minister tells you to.

VI. America is the best country to live in.
1. It is just to all the world.
2. It has wonderful wealth.
3. Its people are intelligent.
4. It is easy to make a good living in America.
5. Americans are respected by others.

VII. We should do nothing to injure others.
1. Our school books tell us to be kind to everybody.
2. Kindness makes other people happy.
3. We wish others to respect our rights.
4. We don't want to be called selfish.
5. Injuring others is sure to get us into trouble.

VIII. When you have a contagious disease, you should stay at home.
1. By so doing you will not expose others.
2. You are sure to get well sooner.
3. You will obey the regulations of the Board of Health.
4. You will be criticised if you go out.
5. Your doctor's bill will be less in the end.

IX. Doctors should be well paid.
1. They spend a long time in getting an education.
2. They work long hours.
3. They are intelligent men.
4. They are of great service to others.
5. Their profession is considered a good one by all people.

X. Lincoln is an example for all to follow.
1. He educated himself.
2. He has a leading place in history.
3. He had charity toward all and malice toward none.
4. He became President of the United States.
5. He had great wisdom.

XI. You should go to church.
1. It is a good way to begin the week.
2. It makes you kinder to other people.
3. You meet many good people.
4. The minister tells you many important things.
5. It makes you familiar with the Bible.

XII. To eat more than one needs is wrong.
1. It deprives others of what they need.
2. The government urges us to save food.
3. Food is expensive.
4. Over-eating injures our health.
5. It may make us gluttons.

In all school grades tested the children on the whole considered the moral reason the best reason, though the difference in favor of the moral reason is not great in the fourth grade. It, however, increased constantly and decidedly through the grammar grades, the high school, and among college students. Whatever may have been true of the conduct of the children, it was quite evident that their judgment with regard to a moral situation became increasingly accurate as they advanced in years and experience.

Besides investigating the moral judgment of children, Liao studied their intellectual honesty. Those tested were given a vo-

cabulary of fifty words arranged in order of difficulty, with instruction to check off all the words they knew sufficiently well to use in a sentence or to define. After the pupils had checked the words they were required to define the last ten that they had checked. It was found that there was a wide variation in the number of words thus checked and also in the number correctly defined or used. There was a fairly high positive relation between the intellectual honesty of the pupil and his school record, but none between the number of the words checked and his care in checking them.

V. Summary

In the foregoing pages the attempt has been made to explain and define the term "general intelligence" as it is commonly used in the field of mental testing, and to show how it is possible to measure innate intelligence—also in this connection to point out certain misunderstandings and dangers involved in the attempt to determine the innate intelligence of an individual or group of individuals. Further, a general sketch of the origin and growth of tests to measure intelligence, culminating in the present group tests for intelligence, has been presented. Particularly, in this connection the general characteristics and forms of intelligence tests have been indicated. Finally, the fact has been emphasized that intelligence tests alone are not sufficient to show the probable efficiency of an individual or his success in school or in life, since character as well as intelligence is a vital element in such success or failure. A brief outline of the work so far done in character testing has been added. In conclusion the following summary of the most important points included in the above discussion may be helpful.

1. The term "general intelligence" signifies an innate capacity or group of related capacities to acquire intelligence in specific situations of life. It can be identified closely with learning ability.

2. This ability is measured by determining the relative degree to which a group of individuals, or a single individual in comparison with a group whose attainment has already been measured, succeed in their scores in tests constructed in such a way that the materials used are of common knowledge and common interest to those so tested.

3. Little or no value can be attached to the results of tests in which the individuals tested vary in any marked degree as to their opportunity and desire to become familiar with the materials of the test employed. Hence children of different social and economic status may score quite differently in such tests not because of any real difference in native intelligence but because of such differences in home surroundings that some are favored while others are handicapped, particularly as far as use of the English language is concerned. Also boys and girls, because of their different interests in the world about them may make quite different average scores in tests as a whole or in various elements included in tests, without differing essentially in native capacities.

4. Intelligence tests thus measure not only native intelligence but interest as well, and to a certain extent character qualities, since learning involves not only intelligence and interest, but also earnestness of purpose and will-to-do.

5. The pioneer in intelligence testing was the French psychologist Binet who, with the assistance of the French physician Simon, drew up the first set of intelligence tests. This was done with a view of determining the number of feeble-minded children in the schools of Paris and segregating them for the purpose of instruction. The Binet tests have since his time been extensively revised, particularly in America, and used for the purpose of testing normal children as well as those of subnormal intelligence. The most extensive revision of these tests has resulted from the work of Terman in California, who has compiled the Stanford-Binet series.

6. Binet, in his scale has a group of tests for each age and a child's intelligence is expressed by indicating the distance along this scale which he can go, thus determining his mental age, and then comparing this mental age with his chronological age (age in years). If his mental age and his chronological age are the same, he is of normal intelligence. If his chronological age is considerably greater than his mental, then he is subnormal. If, however, his mental age is considerably in excess of his chronological age, he is supernormal.

7. In the Stanford scale the mentality of the child is expressed by his I. Q. (intelligence quotient), secured by dividing his

ascertained mental age by his chronological age. Unity means normality; a decimal considerably below unity means subnormality; one considerably above indicates superior intelligence.[20] While the mental age of the child may be used in comparison with his chronological age for purposes of classification, the I. Q. alone cannot be thus used, since children of the same I. Q. may vary greatly in their mental ages as well as in their chronological ages. In classifying children according to I. Q.'s both mental and chronological age must be taken into account.

8. While the Binet tests and their revisions are the most reliable measures of the intelligence of a child that we possess, they require a large amount of time in their use (since they are individual tests), and considerable skill in the technique of administering. The group intelligence tests, which have been developed since 1917, can be given in very much less time, since many children can be tested at a single sitting, and since these tests require much less skill in their administration than do the Binet tests. Therefore, when considerable numbers of children are to be tested, the group tests may legitimately be used. However, when there is doubt in individual cases, some form of the Binet test or individual performance tests should be provided. These usually give more accurate measures than do the group tests. The latter are advantageously used for gross results; the former, for finer distinctions.

9. Finally, it should be remembered in all cases of mental testing that the employment of these tests is merely a means to an end, not an end in itself. Mental tests furnish a certain amount of valuable data, which, when used in connection with other information, such as school attainment, opinions of teachers in regard to children's interests, mentality, and the like, are helpful in classifying pupils in various grades and subjects, in giving them educational advice and direction, and in understanding them as individuals rather than as mere representatives of a group. Administered in a mechanical way and not supplemented by the personal touch, they are often of little value and may be even positively harmful.

---

[20]In practice the I. Q. is usually obtained by multiplying the obtained quotient by 100.

CHAPTER III

STATISTICAL METHODS APPLIED TO EDUCATIONAL TESTING

HAROLD RUGG
The Lincoln School of Teachers College, New York City

The purpose of this chapter is threefold: first, to describe for teachers and administrators common and elementary methods of treating test data (Section I); second, to summarize the newer and more elaborate statistical methods for research workers (Section II); third, to present an annotated bibliography which will put the advanced student of educational statistics in touch with the new methods (Section III).

SECTION I.—ELEMENTARY METHODS OF TREATING TEST DATA[1]

I. SOME IMPORTANT STATISTICAL FACTS

If you give an intelligence test to several hundred school children and draw a graph of your results you will arrive at a figure something like Diagram I-1.

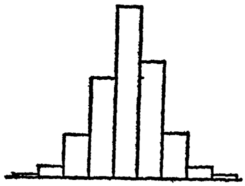
If you give a reading test, say the Burgess Test, your figure will closely resemble Diagram I-2.

If now you should test your pupil's ability to add (or subtract, or multiply, or divide, or to do algebra problems), you would obtain a graph that would look something like Diagram I-3.

In the same way if you should measure any physical trait like stature, or weight, or strength of grip, or girth of chest, or length of forearm, or foot, or what-not, you would arrive at a graph which would look something like Diagram I-4.

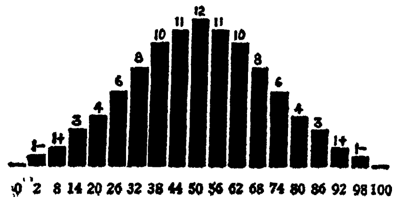
You have now seen four graphs which are typical of the traits with which the school commonly deals. There are three significant

[1] Section I is based upon a forthcoming *Primer of Statistics for Teachers*, to be published by Houghton Mifflin Co., author's copyright.

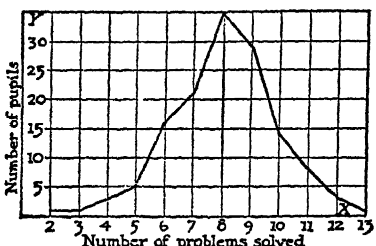


How 904 children distributed in intelligence from very bright to very dull.

No. 1

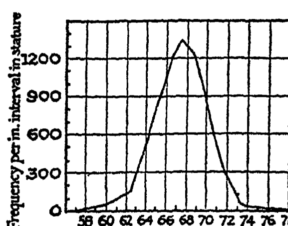


No. 2



Number of pupils who solved various numbers of algebra problems.

No. 3



Heights of 8585 men grouped in inch intervals.

No. 4

Diagram I.—Typical Examples of Distribution of Mental and Physical Traits

facts about these distributions which are at the basis of the schoolman's use of statistical methods:

1. Children vary widely in ability.
2. Graphs of their ability show the same general shape.
3. A large proportion of their abilities cluster so closely around a given value that it typifies the "central tendency" of all.

1. *School children vary widely in ability.* In recent years, however, school people have improved their methods of *measuring* pupil's abilities. Instead of "judging" them, "marking" them on a purely subjective basis, they are carefully testing their abilities to do certain standard tasks. The difficulties of the tasks (examples in arithmetic, words to be spelled, passages to be read, or whatnot) have been carefully determined, by having them worked by

thousands of children. Thus, since the tests are arranged on the basis of known difficulties, and since the tests have been given to large numbers of children, we speak of them as "standardized."

So, charts of pupils' abilities like those in these diagrams are very significant. They show wide differences in such physical traits as stature, in such muscular skills as handwriting; and in various mental abilities like ability to read silently, to solve problems in physics, algebra, etc.

Notice the differences in the *range of ability* in the different traits. In stature, the range of differences is relatively small, although apparently great, 57 inches to 77 inches. In handwriting, in reading, algebra, arithmetic and such subjects, the extreme differences are very much larger. The best pupils do 6 to 12 times as well as the poorest. One can find in a third-grade reading class of 30 pupils, some who read as slowly as 30 words per minute, and others who read as rapidly as 360 words per minute—12 times as fast.

We need not multiply cases. Schoolmen are agreed on this outstanding fact: children whom we have tried to teach in the same section vary widely in ability. Administrators are asking frankly whether it is not futile to try to fit one course of study and one kind of machinery to such gross differences in capacity.

2. *Graphs of pupils' abilities are of much the same shape.* Notice the similarity in shape in all of these graphs, how the curve is very high at the middle and low near each end, how it shades off at the same rate on each side; in other words, how the mediocre pupils are most frequent and the exceptional are less and less frequent, how the very unusual are few and far between.

The *shape* of the graph is very important. It shows how abilities distribute between the very large differences to which we have referred in the preceding section. About one hundred years ago people began collecting physical measurements of human beings. They measured the stature of thousands of men. They measured the circumference and breadth of heads, the length of forearm, weight, chest expansion, and many other anthropometrical traits.

Later when psychological laboratories developed, mental measurements were taken. Not so many cases could be gathered, but

yet enough to give helpful results. Again the recurrence of the same characteristics in the distribution—the piling up of measurements of "mediocrity," the greater and greater infrequency of "unusual people."

3. *Distributions of measurements of intelligence show a "Central Tendency" which is typical of all the measurements.* This is the third striking fact about the abilities of school children. Study the typical figures in Diagram I. Although people vary widely, it is significant that the great mass are much alike. One might generalize from what he finds in the vast accumulation of scientific measurements and from his practical school experience something as follows:

Pupils in school tend to group themselves in a large central mediocrity, flanked on either side by a small but important group of superior and inferior ability. Occasionally one finds exceptional children, brilliant or stupid. These are relatively rare. It is this large, rather compact mediocrity that leads us to speak of the "central tendency" of a distribution.

II. How to Represent School Statistics by Frequency Tables

When you have tested the intelligence or some specific ability of pupils your first task is to set up the data so that the reader can understand them. There are two ways to do this. The clearest way

Table I

Pupils	No. ex. right in 3 minutes	Pupils	No. ex. right in 3 minutes
Adams, Ada	17	Lanterman, Anne	16
Albright, J. H.	11	Lowenthal, Louis	15
Bass, Dan	13	Manning, Fred	10
Brownell, Bessie	10	Marston, Mary	11
Carlson, Anna	18	McMurray, Mabel	12
Crowther, Jas.	4	Mendenhall, Carl	15
Dawes, Janette	9	Metz, Pauline	14
Evans, Isabel	11	Owens, Edward	12
Finch, Geo.	12	Ranney, Geo.	5
Ford, Wm.	11	Reed, Katherine	3
Harris, David	9	Smith, John	14
Herrick, H. E.	8	Wright, Evelyn	13
Hogan, John	6	Wright, Betty	11
Johnson, Emma	19		

is to make a graph. To do that it is necessary to make a table of the data—that is, to have the numbers arranged in some orderly fashion.

You wish to report the scores that your pupils made so that some one else can clearly understand them. You first make a tabulation of the scores with the names of the pupils. The data might appear something like those in Table I. These are not clearly arranged. Your reader wants to know how many made 3, 5, 10, 12, 16, 18, etc. He wants a compact summary with the scores arranged from largest to smallest and with the number of pupils given who made each score.

So you make a *Frequency Table,* and it looks like Table II.

TABLE II

Test Scores Made by Pupils	Number of Pupils Who Made Each Score
19	1
18	1
17	1
16	1
15	2
14	2
13	2
12	3
11	5
10	2
9	2
8	1
7	0
6	1
5	1
4	1
3	1
2	0
	N = 27

HOW TO PLOT A FREQUENCY DIAGRAM

Now to graph the data of the frequency table keep in mind these simple rules:

First: Draw a horizontal line (line OX in Diagram II) and lay out on it the units of the distribution 1, 2, 3, etc. These units are in terms of scores made on the tests. Place the points as far apart

as you can and yet get them all on the paper. This line is the *scale* of measurements of the trait or the fact you are considering.

Second: Draw a vertical line (like OY in Diagram II) through the extreme left end of the horizontal line. Divide this line into a number of units. Remember you are going to represent by *vertical distances* above the horizontal base-line the *number* of individuals or cases. So, to tell how far apart to put your points, find the largest number of cases in the *frequency* column of the table and fit the number of cases to the number of squares that you have vertically above your horizontal base-line. It is better to make the graph steep like Diagrams I-1 to I-4.

Third: Having the units laid off on each line, plot the number of cases by locating points on the cross-section paper above the appropriate points on the base-line. Diagram II shows how it is done for the data of Table 2. Connect the points. This gives a picture or graph of the data. This is sometimes called a *frequency polygon,* or *line-graph.*

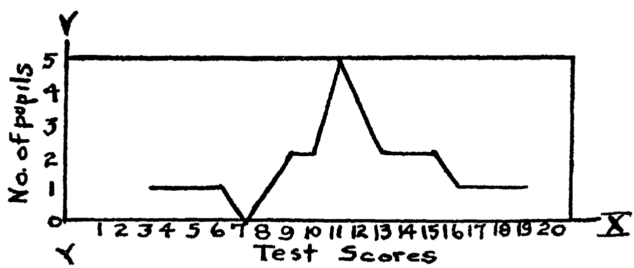


DIAGRAM II

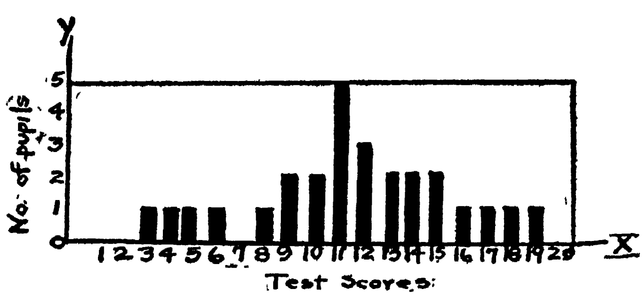


DIAGRAM III

The Bar Graph. It is clearer to some persons to use a *bar graph.* To do this, merely draw a vertical line at each point on the base-line long enough to represent, to scale, the number of cases at that point. These can be made to stand out a little more clearly if the lines are widened to make columns. Still more so if they are blackened like Diagram III.

How Single "Average" Values Helpfully Describe Distributions of Data

Study the distributions in Diagram I. Notice how the cases distinctly concentrate near the middle of the scale. This hump in the graph—this bunching of measures—enables us to describe distributions very easily. We could say, from Diagram I-2, that the "middle half" of the pupils read between 38 and 62, or from Diagram I-3, that mediocre pupils solve from 6 to 10 problems in algebra in five minutes. That is, we can pick out the middle groups in our distributions and tell what they did on our tests.

But this is awkward. We have to use two or three numbers to picture any one group. What we really need is a single number to describe the group. It very frequently happens that we wish to compare two distributions of test scores (*e.g.,* from different classes or schools or school systems) or of school marks, or some other measures of children. We have already studied the first method of summarizing and of comparing such data—preparing a frequency table and a frequency graph. But the simplest types to pick out and compare are the "averages."

The "average" partially describes the distribution. It is a single measure which stands for the central tendency of the data. Let us study a case. Two classes were tested with an algebra test. Diagram IV presents the data as a bar-diagram. Which class is the better? What is the general tendency of the achievements in the two classes? Is the "Central Tendency" of one class better than that of the other? What does "Central Tendency" mean to you? Does it mean the general "feeling" that you have that the bunching of the measures in Miss H's class occurs near a lower score than that in Mr. D's class? That is the sense in which it is

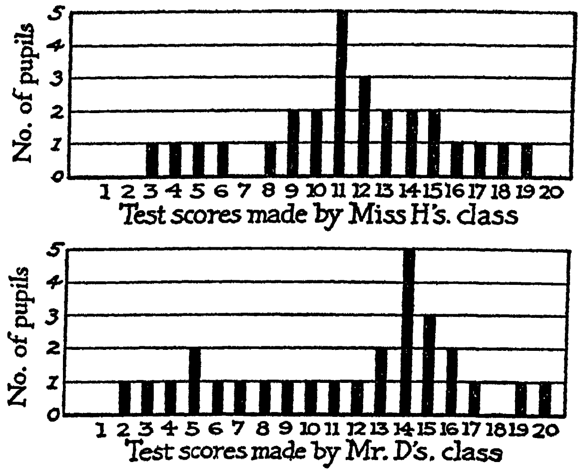


DIAGRAM IV.—GRAPHIC COMPARISON OF SCORES MADE BY TWO CLASSES IN A MATHEMATICS TEST

used by statistical workers to describe the concentration of measures at or around a particular value.

Note how much more definite the comparison of achievement can be made by means of some single average value in each distribution. See how, in Diagram IV, the cases concentrate so decidedly about 11 in one class and 14 in the other that the single central values 11 and 14 describe rather well the central tendencies of the two distributions. Instead of depending on a general feeling of concentration of measures we refer to a single middle or average number which is most typical of the concentration.

There are three such "average" values which are commonly used to describe distributions: (1) the mode, or commonest measure; (2) the median, or middle measure; (3) the arithmetic mean, commonly called the "average."

1. *The Mode: The Commonest Measure.* What is the most conspicuous feature of the various distributions we are comparing? The tall bars in Diagram IV? The high point on the curve in Diagrams I-1 to I-4? What does the extreme height mean? It means "the greatest frequency." The value which occurs most

frequently is called *the mode—la mode,* "the fashion." The modes of Diagrams I-2, I-3, and I-4 are respectively 50, 8, and 67. The mode of Miss H's class is 11, that of Mr. D's class is 14.

Remember that *the mode is the value that occurs most frequently.*

2. *The Median: The Middle Measure.* The median is another average value that is easily determined and that "stands for" all of the measures in the list rather well. It is easiest to *think of the median of a distribution as the middle measure,* and this is sufficiently accurate for practical interpretations.

a. When there is an odd number of cases. For example, if you had a distribution of 11 measures, the median could be thought of as the value of the sixth measure. The approximate median for Miss H's class is 11 because there are 27 measures and the approximate median is the value of the 14th measure.

b. When there is an even number of cases. There is here no middle number. In such an instance the median is taken as the value midway between the values of the two middle cases. Thus, the simple rule is to *find the value of the middle case or the value halfway between the two middle cases.*

c. When the data are so frequent and the values so different that they have to be grouped. Study Table 3. No single middle measure stands out; neither can one distinguish any two middle measures. Sixty-eight measures were so closely of the same value (ranging from 90 to 100), that to economize time and labor, they were grouped together in one interval of 10 units. For very rough purposes you might call the midpoint of the interval the median. In most cases your interpretation of the data would not be different by this method from what it would be were you to compute the median very precisely.

However, the precise computation is not difficult. It consists of finding the value on the scale that exactly cuts the data in two equal parts. In Table 3 there are shown 373 cases. Half of these, 186.5, fall on each side of the median. To locate the median, count the number of cases (up the scale or down) to find the interval which includes the value that divides the distribution in two Thus, in Table 3, counting, say, up from 20.0–29.99 at the end of the

table we get: $8 + 10 + 17 + 21 + 33 + 38 + 54 = 181$. Since there are 68 cases in the next interval and these added to 181 make more than half (186.5) we know the median is somewhere in this interval. Exactly where? To find out, we assume that the 68 cases are evenly distributed over the 10 units of the interval. Then

TABLE III.—TO ILLUSTRATE THE COMPUTATION OF THE MEDIAN

Intelligence Scores	f	
150 - 159.99	6	
140 - 149.99	9	
130 - 139.99	12	
120 - 129.99	25	
110 - 119.99	30	
100 - 109.99	42	
90 - 99.99	68	Md = 90.81
80 - 89.99	54	
70 - 79.99	38	
60 - 69.99	33	
50 - 59.99	21	
40 - 49.99	17	
30 - 39.99	10	
20 - 29.99	8	
	N = 373	

the middle point is evidently $\frac{5.5}{68}$ of the way up that interval. It is located at a point $\frac{5.5}{68} \times 10$ units above 90; that is, at 90.81.

Check. If you count down instead of up, of course, you get the same result. That is $6 + 9 + 12 + 25 + 30 + 42 = 124$. We need $\frac{62.5}{68}$ of the interval 90—100 to locate the median value. Hence $\frac{62.5}{68} \times 10 = 9.19$, and this subtracted from 100 gives 90.81.

The steps involved in computing the median with grouped measures are, then, these:

1. Divide the total number of measures by 2.
2. Count up (or down) the number of measures included in the class-intervals TO the interval that contains the median.
3. Subtract this number from $\frac{N}{2}$ (half the number of measures).
4. Divide the remainder by the number of measures in the interval which contains the median.

5. Multiply by the number of units in the class interval.

6. Add this number to the value of the lower limit of the interval. Use whole numbers 80.0, 75.0, 70.0, etc., instead of 79.99, 74.99, 69.99, etc. If the counting is done from the upper end, subtract from upper limit of the interval.

3. *The Arithmetic Mean, or "Average."* There is a third measure, better known, but less easily used: the *arithmetic "average."* The technical name for this is *"arithmetic mean."* No doubt it is the value we all have in mind when we say "on the average so and so is true." This is the most familiar average value, because it is the one we have been taught to use in school.

a. The "simple average." In the elementary school we teach children how to compute both the "simple average" and the "weighted average." You will recognize the difference from some examples.

Thus the arithmetic mean of 8 and 4 is 6. The mean of $8+5+2$ is 5. The mean of $7+8+4+3$ is $22 \div 4$, or 5.5. So, we say the arithmetic mean or average is *the sum of the values of the measures divided by the number of measures.* We call this form the simple average; each different value occurs only once.

b. The weighted average. Frequently you will want to compute an average when the different values occur more than once, as in Table IV. This illustrates how the "weighted average" is computed.

The word rule for finding the weighted average is the same as for the simple average: *Divide the sum of the values of all the*

TABLE IV

No. of examples worked	Number of pupils who worked each number of examples, *i.e.*, the "frequency" (f)	Products: The values × the corresponding frequency
17	2	34
16	1	16
15	5	75
14	8	112
13	16	208
12	7	84
11	4	44
10	3	30
9	1	9
	$N = 47$	47)612(13.02

measures by the number of measures. That is, multiply each value (17, 16, 15, etc.) by the number of times it occurred (2, 1, 5, etc.) and divide the total (612) by the number of measures (47). This gives the average, 13.02.

How to compute the average when the data are grouped in class-intervals. The intelligence scores of 373 children in a school were as follows:

TABLE V.—TO ILLUSTRATE THE COMPUTATION OF THE ARITHMETIC AVERAGE

Intelligence Scores	f	Mid Point m	f m
150 – 159.99	6	155	930
140 – 149.99	9	145	1305
130 – 139.99	12	135	1620
120 – 129.99	25	125	3125
110 – 119.99	30	115	3450
100 – 109.99	42	105	4410
90 – 99.99	68	95	6460
80 – 89.99	54	85	4590
70 – 79.99	38	75	2850
60 – 69.99	33	65	2145
50 – 59.99	21	55	1155
40 – 49.99	17	45	765
30 – 39.99	10	35	350
20 – 29.99	8	25	200
	N = 373		373)33355(89.42

How can the average be computed for such a case? The actual values of the scores are hidden within the class-intervals. We have to make an assumption regarding the values of the measures. Each interval, 150–159.99, 140–149.99, 130–139.99, etc., has a mid-value; 155, 145, 135, etc. So, for convenience, *we assume that the value of each measure in an interval is the same as the mid-value of the interval.* Of course, that is not really the case. The ten scores in the interval 120–129.99 are 120, 121, 122, 123, 124, 125, 126, 127, 128, and 129; we *call* each one 125. But this does not change our average much, for the true average of these scores is 124.5. From this point we compute the arithmetic average exactly as we do the ordinary weighted average; that is, we multiply the value of the midpoint of each interval by the number of cases in it, total these products and divide by the total number of cases. Table V illustrates this.

4. *Which average value should be used: mode, median, or mean?* Two questions must be answered: which value describes the entire distribution best? Which value is easiest to compute?

a. Which value describes the distribution best? No one value can completely describe a distribution. This fact is clear about all statistical distributions, no matter how widely scattered or how compact the data are. Look at Diagram I-3. The average is 8. But the highest score was 13, while one pupil made as low a score as 2. Certainly no *one* number can completely typify such a distribution of statistics.

This is not an exceptional case. It is typical. Look at the other distributions. What one number can give a mental picture of the great differences between the extremes of the data? No one number, of course. This should be kept clearly in mind in all statistical work. Yet, we need single numbers or at most, a few numbers, to represent different distributions and to enable us to compare them.

What number will serve us best? The answer to the question depends on an important factor—the way the data are scattered over the scale—that is, the *shape of the distribution.* Now, an important fact is that most educational distributions are very symmetrical in shape. For such symmetrical distributions of data the mode, the median, and the mean doubtless will all be nearly the same value. It is this fact of the close equivalence of the values of the median and mean that leads to the conclusion that (for most distributions of data on human traits) *one average value gives as good a description as the other.* And for the simple reason that they are nearly the same value. But it is also generally recognized that the *mode* is not a desirable average to use in accurate work, because it fluctuates too much with slight changes in data.

b. Which average value is the easier to compute: median or mean? Here the decision is clear and definite. The median is more quickly and easily computed than the arithmetic mean. Hence, for distributions which are reasonably symmetrical, since median and mean describe the distribution equally well, use the median because it is more easily computed.

But many administrative facts do not give symmetrical distributions, for example, the distribution of salaries of teachers, ages of pupils, attendance of pupils, receipts and expenditures of school systems. Practically no distribution of this type of facts is symmetrical. Which average would then be the better one? We can answer it by answering the question: Which one describes the data in the entire distribution the better? If accurate comparisons are being made, it is better to use both mean and median.

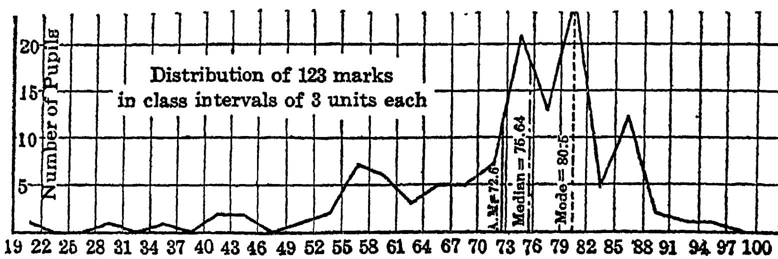


DIAGRAM V.—TO ILLUSTRATE THE COMPARISON OF THE AVERAGE AND THE MEDIAN WITH A SKEWED DISTRIBUTION

For some kinds of distributions the median perhaps sums up the situation better than the mean: for example, a distribution with a long tail containing a few measures of extremely low value (see Diagram V). In computing the arithmetic mean, one high value offsets several of the middle or average values. In computing the median, however, all values count equally. In such distributions, therefore, the median probably gives a better measure of type.

MEASURING THE SCATTERING OF DATA: VARIABILITY

An average does not completely describe a distribution of data. It merely tells about where the middle values are. In the case of distributions of measures of human traits it tells where the measures tend to concentrate; what values occur most frequently. It locates the hump on the curve. It does not tell how wide the hump is—how much the measures are scattered about or away from the average. And it is important to know this. It is the scattering of the mass

we are interested to measure statistically. And there is a very plain way to measure it; namely, to take some convenient fraction of all our measures and state within what values on the scale these are included. The easist number to use is the middle half of the measures, or one of the middle quarters.

Suppose I measured the heights of 8,585 men and found the average height to be 67.46 inches. You would then know one fact about the measurements. This would not tell you anything about the spreading out of the measures. Next suppose I said: "two were as tall as 77 inches, and 3 as short as 57 inches." Now you know two facts, the *average* and the *range*. You know the mean and the extremes. Still you would not know much about the concentration of the measures.

Next, suppose I added that the middle half of the heights (the middle 4292) fell between 65.9 and 69.0. You would know now, that one of the middle quarters (2146) fell between 67.4 and 69 inches, and that the other fell between 65.5 and 67.4. Also that 2146 fell in the eight inches from 69 to 77, and that 2146 more fell in the eight inches from 57 to 65.9. And you would know without seeing the whole distribution, that the measures were decidedly concentrated about the average 67.4 inches.

However, the very clearest way to portray *variability* is to give the graph of the distribution together with some statistical measures. In Diagram VI the whole situation is presented; the average, the range, and the concentration, as shown by the two middle quarters.

Now it is awkward to use the entire phrase "the middle fifty percent falls between." So we use two different symbols to stand for it. The easier one to remember is Q (for quartile). Q is half the difference between the values that take in the middle 50 percent of the cases. In Diagram I-4 the middle fifty percent fall between 65 and 69 inches. That is 2Q is 69-65, or 4 inches. Hence, Q is 2 inches.

There is another symbol for this measure of the middle values: P.E., which stands for *Probable Error*. Q or P.E. may mean the same thing—"the distance on the scale both above and below the

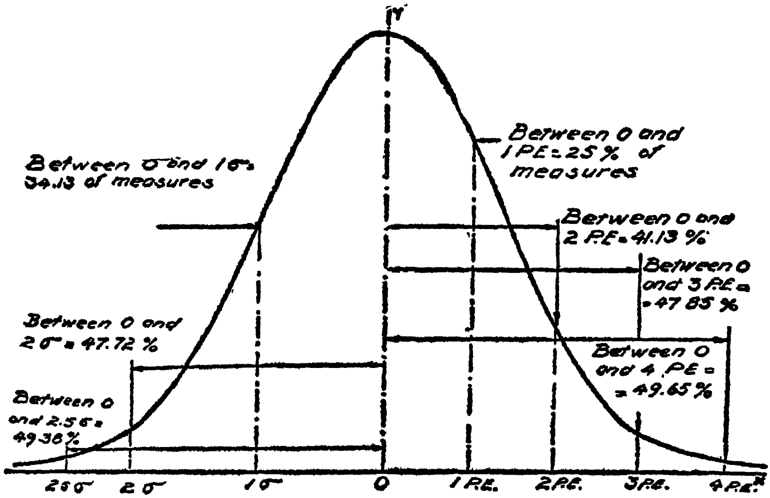


Diagram VI.—To Illustrate the Use of "Standard Deviation," σ, and "Probable Error," *P. E.*, as "Unit Distances on the Scale" (*i.e.*, as Measures of Variability) of a "Normal Frequency Curve"

average that includes 25 percent of the cases." This is strictly true only when the distribution is absolutely normal, or symmetrical.

How to compute Q. Think of any distribution as divided into a number of parts, first, say, halves. The median (Md) is the point on the scale which so divides it. Remember, it is the number of measures you are dividing, not the scale itself.

Next, think of the measures in the distribution as divided into quarters. For example, take the distribution of Diagram VI. That distribution is divided into quarters, not by dividing the units on the base-line into quarters, but by counting in from the largest or from the smallest value until one fourth of the measures, two-fourths, and three-fourths are included. The values on the scale are the quarter points. We call them Q_3 and Q_2 and Q_1. Half the difference (or distance) between Q_3 and Q_1 is Q.

When the measures are grouped in a frequency distribution, determine the quartile points exactly the same way as for the median.

Another Way to Describe Variability: Averaging Deviations from the Mean

There is another convenient way to tell how the measures of a distribution spread out. That is, to picture the amount that the measures as a whole differ from their average.

Look at Table VI. Each measure can be thought of as differing or "deviating" from the average (either mean or median) of the whole distribution. The average is a convenient central point to take because it fluctuates so little. In Table VI the approximate median is 10. Each of the ten measures of value 11 has a "deviation of 1." Each of the four measures of value 14 has a deviation of 4. Similarly the measures of each of the 8 cases of value 9 have a deviation of —1; each of the 5 of value 8 a deviation of —2; and those of value 7 a deviation of —3, etc.

Now the best way to picture these deviations as a whole is to average them disregarding signs. Table VI shows how this is done. (The approximate median 10.0 is used instead of the true median, 10.88, in this illustration.)

TABLE VI

Values	Frequency f	Deviations d	Frequency Deviation fd
17.	1	7	7
16.	0	6	0
15.	3	5	15
14.	4	4	16
13.	5	3	15
12.	7	2	14
11.	10	1	10
10. (approx. md.)	12	0	—
9	8	1	8
8.	5	2	10
7.	3	3	9
6.	1	4	4
5.	2	5	10
4.	1	6	6
3.	1	7	7
	63		131

$131 \div 63 = 2.07$, the Average Deviation, A.D.

Another Way to Describe Variability: The Standard Deviation

Perhaps you will find it more helpful to think of distributions as divided into thirds, instead of halves or quarters. If so, the *standard deviation* will be clear to you as a measure of variability. In round numbers it is the difference in value from the average that includes one-third of the entire number of cases. Diagram VI illustrates this measure.

This deviation, the standard deviation, is used a great deal in accurate statistical work and its symbol is S.D., or oftener σ (sigma). Between the mean and -1σ on the left side about one-third of all the measures are included. Accurately, on a particular distribution known as "normal," 68.26 percent of the measures are taken in between 1σ and -1σ.

For practical interpretive purposes, Q, P.E. and A.D. may each be thought of as taking in about one-fourth of the measures on each side of the average, and σ as taking in one third.

How to Compute the Standard Deviation[1]

The standard deviation is computed much like the average deviation. The chief difference is that each "deviation" is squared and the square root of the average is taken.

Table VII illustrates the method. In it 477 is divided by 63 and the square root of the quotient gives 2.75, the standard deviation.

How to Compare the Variability of Distributions of Data

One method of telling when one distribution is larger than another is to compare the averages. Differences between the distributions may consist, however, not in average value, but in the scattering of the measures, in the variability. The question will arise: Can we tell which of two distributions is the more variable by comparing two Q's or two A.D.'s or two S.D.'s? Only under two conditions: first, the units of measurement must be the same; second,

[1] A short method of computing the standard deviation for grouped data will be found in the writer's *Statistical Methods Applied to Education*, p. 163.

the average values must be approximately the same. Under these conditions the size of the two Q's or the two A.D.'s or the two S.D.'s will tell you the relative variability of the two distributions.

TABLE VII.—TO ILLUSTRATE THE COMPUTATION OF THE STANDARD DEVIATION

	f	d	fd	fd^2
17	1	7	7	49
16	0	6	0	0
15	3	5	15	75
14	4	4	16	64
13	5	3	15	45
12	7	2	14	28
11	10	1	10	10
10	12	0	0	
9	8	–1	– 8	8
8	5	–2	–10	20
7	3	–3	– 9	27
6	1	–4	– 4	16
5	2	–5	–10	50
4	1	–6	– 6	36
3	1	–7	– 7	49
	N = 63			477

HOW MANY AND WHICH CASES SHALL WE TEST? IN EDUCATION WE DEAL WITH SAMPLES

A measure of the abilities of the pupils in *one* class would usually give a very irregular shape. Teaching groups are small, generally less than fifty. Single classes may be regarded as "samples." Now, of course, no important generalizations can be made from such samples as these. We would need much larger numbers.

Suppose, for example, we wished to know the "standard" reading ability of third-grade children against which any teacher might check the work of her class. One way to set such a "standard" is to find the average reading ability of third-grade children on a particular test. Another, perhaps, to find the average of the best third, etc.

How many children shall be tested? One class of forty? Three classes in a given elementary building? The third grades of all buildings in a city? All the third-grade children in the country? This is an important statistical question.

Clearly one class is not enough. Comparison of the average reading abilities of two third-grade classes from the same building proves that. How? The averages are different. And a norm or standard, for a given kind of teaching and testing, should be constant. On the other hand, we cannot afford to test all the third-grade children of the country—several million. How many shall we test to be sure that we have the "norm"? The answer comes from the theory of "sampling." As we increase the number of cases, the regularity of the distribution increases. When we have several thousand cases, the polygon made up of straight lines becomes so continuous that it may fairly be called a continuous curve.

Now, when two or more distributions from the same data are very continuous, their averages are always very closely the same. And this known fact gives us the criterion for the size of a representative sample: A representative or *random sample* is such a number of cases that if another sample like it be taken, the averages, the measures of variability, and the distributions themselves are closely the same. We cannot generalize as to the number of cases needed with a given kind of data. That will depend upon the condition of the particular problem. We have already learned, however, that for most facts from education, 500 cases are necessary to give a very continuous distribution. When setting a "norm" for a given trait, however, it would doubtless be necessary to make thousands of measurements. For example, see Diagram I, 4, giving the heights of 8585 men. The average is 67.46 inches. Doubtless the average heights of another 8000 or 9000 men, provided they were selected at random, that is by chance, would not be much different from 67.46. For example, there are statistical methods by which we can predict with practical certainty that the average height of another group of 8585 men, selected in the same way, would be within .08 inches of 67.46 (*i.e.,* within $\pm 4 \times$ P.E. which is .02 inches). A practical way to express our ideas would be to say: "The chances are even that if we took by chance, another sample of 8585 men, the average height would be within .02 inches of 67.46." This .02 inches is called the Probable Error (P.E.) of the average.

It is of great importance to be able to make such predictions with certainty. It tells us rather definitely whether we ought to enlarge our number of cases. In the case of the average height of men, if the uses we were making of the data demanded no greater precision in the average than .08 of an inch, then 8585 is certainly a large enough number of cases. For some uses a much smaller number of cases would be satisfactory.

How to Tell Whether Two Things Are Related: Correlation

Do the pupils who read most rapidly comprehend best what they read? Are those who do the formal arithmetical processes skillfully the ones who reason best? Are those who know the most facts in geography the ones who "generalize" best about problem situations in geography? Are the most "intelligent," the best spellers? These are rather important pedagogical questions. There are many others like them. We used to dispose of them rather arbitrarily and quite without evidence. We had certain preconceptions about reading ability, for example. Reading to be well done had to be slowly and carefully done. Is it true, though? If we measure pupils' *rates* of reading and also their *comprehension* of what they read, what do we find? Do the slowest readers comprehend best what they read? Not all. Some do and some do not. Diagram VII is one way of showing this. It shows the names of pupils in exact rank order in both rate and comprehension. Each line connects the two rank positions of the same pupil—his rank in the group in rate of reading with his rank in ability to comprehend.

If rate of reading were perfectly related (or "co-related," or "correlated," as we shall call it) to comprehension, then each of the connecting lines would be exactly horizontal. Each pupil would occupy exactly the same rank position in rate and in comprehension. The first in rate would be the first in comprehension; the second in rate would be the second in comprehension; and so on to the last in rate, who would also be last in comprehension. This would be called "perfect correlation." If it obtained, the two traits, "ability to read rapidly" and "ability to comprehend what is read" would be equally developed in people.

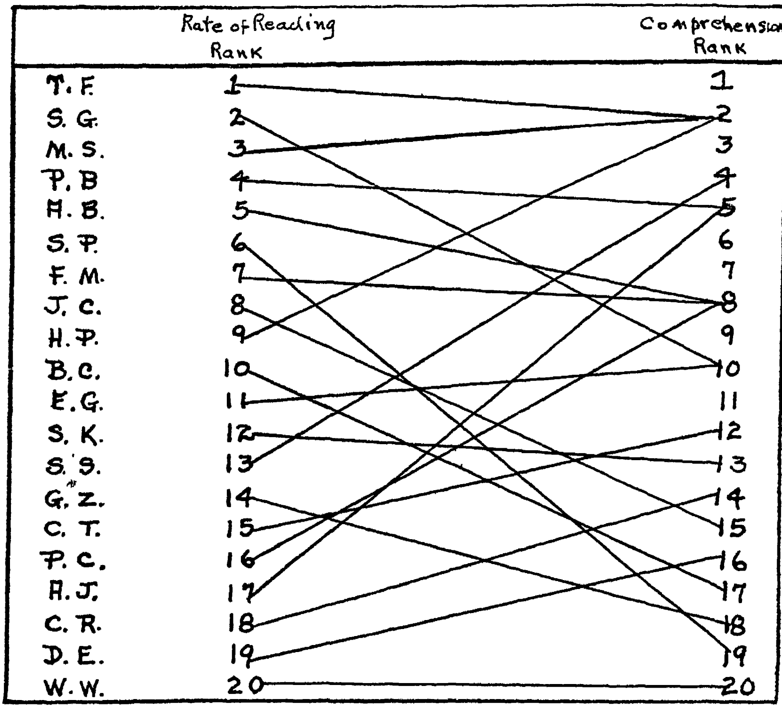


DIAGRAM VII.—COMPARISON OF RELATIVE POSITIONS OF PUPILS IN RATE OF READING AND IN COMPREHENSION

But Diagram VII shows that the two traits are not perfectly correlated. In fact *no two human abilities are perfectly correlated.* The lines tend to be somewhat horizontal. The pupils fall into about the same general division on each scale, but do not occupy exactly the same ranks.

We can tell from Diagram VII only in a general way how closely the two traits are correlated. There are other ways to tell more exactly.

One way is shown by Table VIII. The pupils are grouped in five groups with respect to their ability to comprehend. The average reading rates are then given for each group. The best four pupils read, on the average, nearly three times as fast as the poor-

est four, 252 against 92. And there is a steady increase from the poorest to the best readers, 92, 132, 169, 208, and 252. Such evidence tells us that there is a distinct tendency for those who read rapidly to comprehend best what they read. *Vice versa,* the slowest readers comprehend the least. And the relation appears to hold well throughout the group.

TABLE VIII—ILLUSTRATING HOW ABILITY TO COMPREHEND IS RELATED TO RATE OF READING*

How the pupils were grouped	Scores in comprehension made by the pupils	Rate (words per minute) at which different groups read
The four best in comprehension in the class	98	252
The next four best	86.5	208
The middle four	91.5	169
Four who were inferior in comprehension	91	132
The four poorest in comprehension	82	92

* These pupils were carefully tested by the Courtis and by the Burgess Reading Tests. Their ability to comprehend was marked rather accurately and their rates very accurately.

But this method of telling to what extent things are related is not very exact. It leads only to statements about "tendencies," to "in general it is true," to "there appears to be a correlation," etc. We need more exact methods, so we use single numbers.

The Coefficient of Correlation, "r"

In a perfect correlation each pupil occupies the same position on each scale. We say that the correlation is 100, or better yet 1.0. It is the "highest" we could get. It is inconceivable that two things could be more "highly" or "perfectly" correlated. We call this number the *coefficient of correlation.* The symbol for it is "r." You would read r. .49, as "the coefficient of correlation is .49."

Now suppose the most rapid reader was the poorest reader, the second most rapid reader was the next to poorest in comprehension, the third poorest in rate was the third poorest in comprehension, and so on throughout the entire group. Then we would have "negative" or "inverse" correlation, where the high in one trait are the low in the other. Actually, we know that human traits are not so inversely, or negatively related.

Now this is the most extreme case of "negative" correlation we could have. The first are last and the last are first. We use the number —1.0 to express this extreme *negative* correlation just as +1.0 is used for perfect *positive* correlation. Thus we can think of the correlation (relationship) between the two things as expressed by a single number. And we know now that that number will always be between +1.0 and —1.0. Think of the amount of correlation, the coefficients of correlation as laid out along a scale, like Diagram VIII.

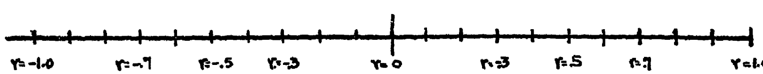


DIAGRAM VIII.—TO ILLUSTRATE HOW "r" MAY VARY FROM — 1 TO + 1

Now, clearly, r can vary all the way from + 1.0 to 0 and from 0 to — 1.0. It can be +.7, or +.02, or +.12 or 0, or —.07, or —.29, or —.82, etc. Is "r =.70" an example of "high" correlation or does r have to be .80 or .90 to be "high"? Some educationists have been very careless in their interpretations of values of r. Some have called r =.25, "distinctly marked correlation" and .40 "high correlation." Others interpret "high" to be anything above .60 and any value of r below .20 as "very low."

By *"high" correlation* is commonly meant a value of r which is about .5 to .7. By "very high" correlation an r which is in the neighborhood of .8 and .9. By "marked" correlation an r ranging from .35 to say .50. By "low" correlation an r about .20 to .35. When r gets as low as .10, it is safe to conclude that there is no significant degree of relationship.

How to Compute the Coefficient of Correlation, "r"

1. The 'Rank' Methods

The easiest way to compute r is to rank each set of measures and use a simple formula:

$$\rho = 1 - \frac{6 \Sigma D^2}{N (N^2 - 1)}$$

In this formula D is the total of the differences between the ranks of the measures in the two series. N is the total number of measures. The steps in the computation of ρ is as follows (see Table IX):

1. Rank the measures in order of size, beginning with the smallest or largest.
2. Subtract the rank of each measure in the first series from its corresponding rank in the second series. Call this *D*, the difference in rank. Tabulate these as positive, negative, or 0.
3. Square each of these differences, giving the column headed D^2.
4. Sum the D^2's giving ΣD^2 or Σg.
5. Multiply ΣD^2 or Σg by 6.
6. Divide $6 \Sigma D^2$ by $N (N^2 - 1)$.
7. Subtract the quotient in either case from 1. This is ρ for the first method, *R* for the second.
8. Transmute ρ into *r* by reading proper value from tables.
   Transmute *R* into *r* by reading proper values from tables.
   This method is called "Spearman's Method of Rank."
   There is a still simpler method: "Spearman's Footrule for Correlation." The formula is:

$$R = 1 - \frac{6 \Sigma g}{N^2 - 1},$$

in which *g* is any positive difference. So the chief distinction between the two methods is that in the first the differences are squared—in the second, not. Either method can be used—probably the squared difference method will be more satisfactory. The writer recommends that rank methods be used only for small numbers of cases, say less than 30 to 40, and especially when the interest is in finding out the correlation for relative position only.

TABLE IX.—TO ILLUSTRATE COMPUTATION OF CORRELATION BY SPEARMAN'S RANK-DIFFERENCE METHOD

	Rank Rate of Reading	Rank Comprehension	D	D^2
T.F.	1	2	1	1
S.G.	2	10.5	8.5	72.25
M.S.	3	2	-1	1
P.B.	4	5.5	1.5	2.25
A.B.	5	8	3	9
S.P.	6	19	13	169
F.M.	7	8	1	1
J.C.	8	15	7	49
H.P.	9	2	-7	49
B.C.	10	17	7	49
E.G.	11	10.5	- .5	.25
S.K.	12	13	1	1
S.S.	13	4	-9	81
G.Z.	14	18	4	16
C.T.	15	12	-3	9
P.C.	16	8	-8	64
A.J.	17	5.5	-11.5	132.25
C.R.	18	14	-4	16
D.E.	19	16	-3	9
W.W.	20	20	0	
$N = 20$				$\Sigma D^2 = 731$

$$\rho = 1 - \frac{6\,\Sigma D^2}{N\,(N^2 - 1)} = 1 - \frac{4386}{7980} = .45$$

$$r = .47$$

2. The Product-Moment Method

It is more common to compute correlation by what is known as Pearson's *product-moment* formula. The simplest form to use is:

$$r = \frac{\Sigma\, x.y}{\sqrt{\Sigma\, x^2 . \Sigma\, y^2}},$$

in which x is the difference between the average of one distribution and any measure in the distribution and y is a like difference for the other distribution.

Table X shows how this is done for the same distributions as before; *i.e.,* for the correlation between rate and comprehension in reading.

TABLE X.—TO ILLUSTRATE COMPUTATION OF THE COEFFICIENT OF CORRELATION BY THE PEARSON PRODUCT-MOMENT METHOD

Pupil	Score in I	Score in II	x diff. of Scores in I from Average	y diff. of Scores in II from Average	x^2	y^2	xy
T.F.	290	100	120	10	14400	100	1200
S.G.	261	94	91	4	8281	16	364
M.S.	230	100	60	10	3600	100	600
P.B.	226	97	56	7	3136	49	392
A.B.	221	96	51	6	2601	36	306
S.P.	211	66	41	–24	1681	576	–984
F.M.	204	96	34	6	1156	36	204
J.C.	196	88	26	–18	676	324	–468
H.P.	194	100	24	10	576	100	240
B.C.	173	81	3	– 9	9	81	– 27
E.G.	156	94	– 14	4	196	16	– 56
S.K.	153	91	– 17	1-	289	1	– 17
S.S.	147	98	– 23	8	529	64	–184
G.Z.	142	76	– 28	–14	784	196	392
C.T.	122	93	– 48	3	2304	9	–144
P.C.	116	96	– 54	6	2916	36	–324
A.J.	110	97	– 60	7	3600	49	–420
C.R.	103	90	– 67	0	4489	0	0
D.E.	94	83	– 76	– 7	5776	49	532
W.W.	62	58	–108	–32	11664	1024	3456
Average=170	*90*				*68663*	*2862*	*5082*

$$r = \frac{\Sigma\, x.y}{\sqrt{\Sigma\, x^2 . \Sigma\, y^2}} = \frac{5082}{\sqrt{68663 \times 2862}}$$

$$= \frac{5082}{14019} = .36$$

$$\text{P.E.} = .6745\ \frac{1 - r^2}{\sqrt{N}} = \pm .13$$

How reliable is the correlation coefficient? If we correlated rate and comprehension in many other classes, would we continue to get $r = .36$ as we did in this one? Or would r vary widely, say from .2 to .8? How can we tell? We might take many classes and compute the r's. This is impracticable. It is possible to get much light from what is known as the probable error of the coefficient. P.E. This is computed from the formula:

$$\text{P.E.}_r = .6745\ \ \frac{1 - r^2}{\sqrt{N}},$$

in which r is the coefficient of correlation and N is the number of cases. In the improvement of methods the computation of coefficients of correlation and of probable errors plays an important part.[2] Diagram VI shows that the probable error is a number that, added to the average and subtracted from it, takes in the middle half of the measures. From Diagram I-4 we found that the average height of 8585 men was 67.4 inches and the P.E. of the distribution 2.0 inches; half the men fell between 65.4 and 69.4 inches. Since 50 percent more fell outside, we say "the chances are even" (1 to 1) that the height of any person selected at random will be between 65.4 and 69.4.

Now study diagram I-4 again. Between $\pm$ 2 P. E., 82.26 percent of the cases are included, and 17.74 percent fall outside. So we say: the chances are about 4.5 to 1 that the height of any person selected at random will be between 63.4 and 71.4 inches (*i.e.*, 67.4 $\pm$ 4 inches).

In the same way, if the P.E. of a correlation coefficient of .50 is, say, .07, it means that the chances that the true value lies within

$\pm$ 1 P.E. are 1:1
$\pm$ 2 P.E. are 4.5:1
$\pm$ 3 P.E. are 21:1
$\pm$ 4 P.E. are 142:1, etc.

To be regarded as sound, we demand that a coefficient of correlation, r, be at least four times as large as its P.E.

We are now determining the probable errors of the scores made by persons on tests. For example, the P.E. of an I.Q. (Stanford-Binet) is about 3.5 points. Otis, who has worked upon the matter says: "An I.Q. is probably in error to the extent of about 6 points or more in a quarter of the cases, 10 points or more in one case in ten, and 14 points or more in one case in a hundred." The P.E. of the mental age of an adult determined by the Stanford-Binet test is about 6 months. "That is, in 50 percent of cases, mental ages of adults may be assumed to be correct within 6 months."

---

[2] See *Statistical Methods Applied to Education*, pp. 233-275.

3. Computing Correlation from "Scatter-Diagrams"

To get the clearest understanding of the correlation between two things, one should plot a "scatter-diagram" of the pairs of measurements, like Diagram IX. The computation can be done by an

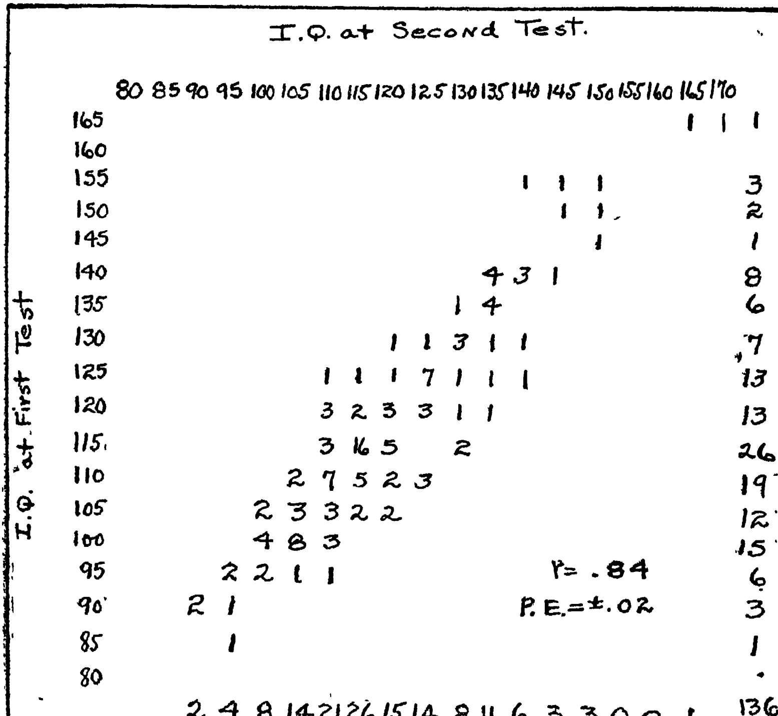


DIAGRAM IX.—TO ILLUSTRATE TABULATION OF PAIRS OF MEASURES FOR COMPUTATION OF CORRELATION BY THE "ASSUMED-MEAN" METHOD (PRODUCT-MOMENT)

abbreviated method.[3] If all the cases occurred in the squares along a diagonal we would have perfect correlation, $r = +1.0$, or -1.0. If the cases were widely scattered over the squares, then r would become small and the correlation would be nearly zero, that is, a "chance" correspondence.

[3] Described in *Statistical Methods Applied to Education.*

SECTION II.—THE DEVELOPMENT OF STATISTICAL METHODS IN EDUCATIONAL RESEARCH, 1916-1921[4]

The preceding pages have been written in the attempt to acquaint school teachers and administrators with common and elementary methods of treating test data. It is the purpose of the remainder of this chapter to bring together for research workers the newer methods employed in the treatment of research material.

Testing Correlation Data for Linearity of Regression

We comment first on the fact that practically no use has been, or is being made of non-linear relationship. The general formula for correlation is strictly applicable to linear relationships only. A non-linear relationship must be reduced to a linear relationship before the formula is applied. Thousands of computations are being made of the correlations between different mental functions. The relationships are so universally linear that practically no reports are made of precaution having been taken to determine the linearity of regression, and it is true that in the case of the correlation between mental traits the case of linearity is becoming more firmly established. It should be pointed out, however, that, as workers in educational research deal more extensively with the correlation of administrative facts, the precaution should be taken to test the linearity of the regression. For example, one of the writers has collected correlations for such things as size of class, and cost of instruction, costs of the several subjects, etc. In these examples no case has been found of straight-line regression. To use the product-moment formula of such variables is to hide the truth. For example, non-linear tables that show an η of .90 frequently give values of r as low as .40 when the product-moment formula is applied.

Two Types of Statistical Procedure Now Employed in Education

The widespread use of mental and educational tests paralleling the establishment of school bureaus of research has stimulated the use of two types of statistical procedure.

---

[4] Cecile Colloton collaborated with the writer in the preparation of this section.

First, bureau directors and school administrators are rapidly becoming familiar with, and are using, the graphic and statistical methods of averages, variability, and correlation. It is not uncommon for the standard methods of determining relationship (referred to in the foregoing sections) to be used by these workers. The elementary uses of probability, and the determining of correlation by more complicated methods are, however, not being taken up by these workers. This probably is for the reason that most of our so-called "educational research" is not research at all. It is largely school administration: the giving of tests, the determining of scores, computations of averages and their comparison with "norms," the occasional study of individual pupils and the making of remedial recommendations. This is the work of the practitioner in diagnosing and prescribing treatment. Naturally, only the most elementary statistical methods are employed, namely, the use of averages, measures of variability. Correlation is only rarely used.

Second, in addition to these administrators, a small nucleus of workers, made up of professional students of education and graduate students in our schools of education, are using more elaborate methods. It is interesting to see the parallelism in the development of the science of education with that of the older established sciences. In education today there is a marked practical demand for a statistical technique by which our educational and mental measuring instruments can be improved. In response to it new methods of determining their reliability are being developed. This is engrossing the attention of many of our students of statistical methods.

We publish at the end of this chapter an annotated bibliography of writings dealing with the recent use of more elaborate statistical methods. It is important to note that the refinement of methods is a product of the past five years. A few of our workers, notably Kelley, Otis, Ruml, Rosenow, Thurstone, were engaged in their first studies in the years 1912-1916. Our entrance into the war postponed the publication of some of this material, *e.g.*, Otis' critical work on the reliability of tests. One more historical comment is worth making in passing: the leadership in development of statistical methods appears to be passing out of the hands of the

student of laboratory and pure psychology (where it was for a generation) into the hands of a younger generation of research workers in the fields of education, industrial psychology and personnel administration.

There have been two distinctive leads in this refinement of statistical methods by educationists: first (perhaps the more engrossing at the present time) is the development of methods to determine the reliability of mental and educational tests; second, an interest pervading education, in common with other social sciences, is the development of statistical methods to predict future conditions, as for example, success in school or in an occupation.

Determination of the Reliability of Tests

Current methods of determining the reliability of tests are fourfold: (1) determination of the agreement of the distribution of the test scores with the known or probable distribution of the trait; (2) determination of the number of times a test would have to be repeated in order to discover "with any desired degree of reliability the relative standing of the pupils" taking the test, *i.e.*, self-correlation; (3) correlation of the test scores with a sound criterion, *i.e.*, with other and independent measures of the trait; (4) determination of the probable errors (or standard deviations) of single test scores.

1. Agreement of the Distribution of the Test Scores with the Known or Probable Distribution of the Trait

This is a necessary step in the construction of a scale and has been employed from the beginning of the movement. Examples appear in the Buckingham Spelling Scale, the Ayres Spelling Scale, the Burgess Reading Scale, etc. Such examples also illustrate the attempt that is being made to improve tests by assuming that that test is the more reliable in which the elements of the test are distributed at equal intervals on the base-line of a distribution curve. The normal probability curve is being employed universally as the best approximation to the shape of the distribution of these total abilities—"reading ability," "handwriting ability," etc. It should

be noted that this whole method of comparison with a law of distribution is an inadequate measure of the *reliability* of the test.

2. Number of Repetitions of the Test Necessary to Secure a Given Reliability; Self-Correlation

A very large amount of work is being done along this line. Most of it consists in determining "coefficients of reliability" by means of Brown's formula.[5] This is the coefficient of reliability $= \dfrac{n\,r}{1 + (n - 1)\,r}$, in which n equals the number of repetitions and r is the coefficient of correlation from two applications of a test. Suppose for illustration, $n = 2$, then coefficient $= \dfrac{2\,r}{1 + r}$ This coefficient of reliability enables one to predict how closely the combined results of any two trials of a single test would correlate with like combined results from two other trials with the same test.[6] Conversely, setting any desired degree of reliability, the formula enables one to predict the number of repetitions necessary.

First Limitation of the Method. Dr. Burgess has pointed out one of the limitations of the use of the formula so well that I shall quote her discussion.

"The coefficients measure the degree to which children who made good scores in the first test also made good ones in the second test, and conversely, the degree to which those who did poorly the first time also did poorly the second time. When the correlations are fairly high, they show that there was substantial agreement in the results of the two testings, but that this fell short of being complete. These results give us more information with regard to the children than they do with regard to the test. They show us that some children who did well on the first day performed quite differently on the following day; and the same type of statement may be made about those who made poor records on the first trial. . . .

"The important fact to remember about such scores is that they may vary from day to day and still be actual true measures of ability on each occasion. Under such conditions the fact that the scores vary from trial to trial does not reflect any inaccuracy and inadequacy of the test or measuring device. . . .

---

[5] Brown, Wm. *The Essentials of Mental Measurement.* Cambridge University Press, London, England, 1911, pp. 101-2.

[6] The best elementary discussion of this is in Burgess, *The Measurement of Silent Reading.* Russell Sage Foundation, pp. 129-133.

"What Brown's formula really does is to compare the coefficient of correlation between one pair of results from two applications of a test with the coefficient of correlation that would be obtained between one average of scores from two or more testings and another similar average of scores from two or more testings. . . .

"The method is of limited value because it is impossible to tell whether the correlation between the first two testings is low, average, or high. In the case of the data given by Professor Thorndike, and referred to in the preceding section, the correlations between the various testings of the same individuals with the same test ranged from .36 to .90. If the coefficient of reliability were based on the lowest correlation it would indicate that the results of no fewer than 16 different testings would have to be amalgamated in order to give a reliability coefficient of .90. If it were based on the highest correlation it would indicate that no amalgamation at all would be necessary to produce the same result."

Second Limitation of the Method. One of the most frequently used methods of determining the reliability of a test is to find its self-correlation, *i.e.,* the correlation of one form of the test with a second form. The second form is to be composed of material like that in the first form, but not identical with it. We have referred to one danger in using coefficients of reliability obtained through self-correlation. There is another, namely that *the size of the coefficient depends upon the spread of the group tested.* The spread of ability in a single school grade is probably not more than one third what it is in 12 grades. This difference in dispersion will change markedly the size of the coefficient. For example, Otis gave the Stanford-Binet test to 180 adult males. He divided the test questions into two halves (or forms) so that the first form contained the first half of the questions for each age-level, and the second form contained the second half. The correlation *for the entire group* was .85. Taking only those individuals whose mental ages fell between 13 and 16:11, the correlation proved to be only .44. Taking only those individuals whose mental ages fell between 13 and 14:11, r was —.14. Taking now only those between ages 13 and 13:11, the correlation was —.62.

Kelley has commented on the same pitfall and has developed a formula by which one can determine, knowing the ratio of the variability in the two groups, what the size of the correlations would have to be, to be comparable. His formula is:

$$\frac{\sigma_t}{\sigma_T} = \frac{\sqrt{r\,(1-R)}}{\sqrt{R\,(1-r)}},$$

in which σ_t and σ_T are the standard deviations of the two groups in terms of true ability and r and R are the reliability coefficients of the two groups. He takes an illustrative case. "To secure a reliability coefficient of 0.40 from a group composed of children in a single grade is probably indicative of greater, not less, reliability than to secure a reliability coefficient of 0.90 from a group composed of children from the second to the twelfth grades." He assumes $\sigma_T = 4\,\sigma_t$ and $r = 0.40$. Solving the above equation gives $R = 0.914$.

If the standard deviations of the scores in the two groups are known, one does not need to make an assumption about dispersion and can use this formula:

$$\frac{\sigma}{\Sigma} = \frac{\sqrt{1-R}}{\sqrt{1-r}},$$

in which σ and Σ are the standard deviations of the two groups.

This equation can be employed to tell whether an increase in a correlation is due to its being found from a particular part of the range. This equation can, therefore, be used as a criterion to tell whether a test is equally effective in a range Σ as in another range σ.

3. Correlation of Test Scores with a Criterion

Correlation of test scores with a criterion is primarily a measure of validity, not of reliability. Kelley has commented on the fact that "if a measure correlates very highly with known measures of capacity, it must of necessity have a fair degree of reliability, but, as the converse is not true—that if a test has high reliability, it will correlate well with a valid criterion—correlation with a good criterion should be used as a measure of validity and not of reliability." Now it is very important to know the *validity* of a test, that is, whether it measures what it purports to measure. But we should not confuse *what* traits our tests measure with *how well* they measure them. Nevertheless, Kelley shows that in order to determine both what a test measures and how well it measures it,

we must know (1) the correlation of test with criterion, (2) the reliability of the test, (3) the reliability of the criterion. The difficulties which we now face in improving our tests are shown by the fact that the reliability of the criterion is rarely known and that we have not carried far as yet the determination of the reliability of our tests. For illustrations the reader should see Kelley's article "The Reliability of Test Scores" (see Section III, Bibliography, Ref. 1).

4. Determination of the Probable Errors of Test Scores

This lead appears to give the greatest promise of helpful results, and considerable application is made of it. It is now postulated that that test is the more reliable which gives the smaller probable errors in individual scores. Care is taken to see that probable errors are expressed (using Kelley's terminology) either (1) in terms of a measure of deviation of the group tested, or (2) in terms of the deviation of some standardized group, say "unselected English-speaking 12-year-olds," or (3) in terms of the difference between two standardized groups, say "unselected children of two different ages."

One of the best examples of this method of determining reliability is the work that is being done on the Stanford-Binet test. A number of individuals have worked upon it. It is now possible to say that the P.E. of an I.Q. is approximately constant and is about 3.5 points (Ref. 2).

The chief use of probable errors is in connection with the need to estimate true (average) test scores from known (single) test scores. (Remember that the "true" score is the average of the many scores that individuals would make if tested under like conditions on a large number of forms of the test.) The most easily interpreted formula to use is that for the probable error of estimate:

$$\text{P.E. est.} = .6745\,\sigma\sqrt{1 - r^2}$$

There is a very real disadvantage to using the smallness of probable errors of estimate, namely, that if the units of two tests (say for reading, or spelling, etc.) are different, the P.E.'s cannot be compared unless the units are equated in some fashion. For

that reason Kelley has proposed that we define our standard groups so that another investigator can duplicate them, *e.g.*, take "unselected English-speaking 12-year-olds." He has also proposed that the difference between the mean scores of unselected 12-year-olds and 13-year-olds be taken as the unit and that the probable error of estimate of tests be expressed in terms of this unit. There are so many other complicating factors (*e.g.*, inequality in rate of growth) that it should be held in mind that these are merely suggestions to stimulate thought and discussion.

Development of Methods of Scientific Analysis and Prediction

1. Multiple Correlation and Partial Regression Equations

The primary purpose of science is the discovery of law and the bases of prediction. We devote ourselves to their study only that we may control both our conduct and our environment. There is no clearer evidence that education is becoming a science than the spectacular manner of its adoption of the methods of statistical correlation, especially the theory and practice of multiple correlation. The annotated bibliography at the end of this chapter provides a striking exhibit of the rapidity with which our great social sciences are assuming their scientific obligations.

Probably no better illustration can be found of the possibility of using multiple correlation to control our social and economic environment than Moore's recent use of it (1917) to forecast the yield and price of cotton. He has shown that if the rainfall and temperature, four, three, and two months, respectively, in advance of the harvest are known, one can predict the yield of the cotton crop with (1) a multiple regression equation (either of three or four variables) of the type:

$$x_0 = b_1 x_1 + b_2 x_2$$

where x_0 is the unknown yield, x_1 is the known data of rainfall, and x_2 the known data of sunshine; (2) by calculating the degree of relationship between these variables by the coefficient of multiple correlation:

$$R = \frac{\sqrt{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{13}\, r_{23}}}{\sqrt{1 - r^2_{23}}};$$

(3) by determining the accuracy of the multiple regression equation as a forecasting formula by calculating the standard error of estimate:

$$S = \sigma_0 \sqrt{1 - R^2}$$

He shows that prediction by the use of multiple correlation is more accurate than the official forecasts of the Federal Department of Agriculture with its wonderful statistical organization. "By a connection with many thousands of correspondents, by field-agents, by special experts in crop estimates, by a Bureau of Statistics and a Crop-Reporting Board, information has been systematically gathered and tabulated, and for several decades monthly reports have been issued throughout the growth season of the crop. Extraordinary precautions have been taken to prevent any leakage of the precious information before it is given to the public." Thus, in a field where natural causes dominate, fundamental causal connections can be, and are being discovered by multiple correlation. Likewise in the field of social causes.

Although it is the infant of the sciences, education has made a most important beginning in prediction by multiple correlation. One outstanding use is being made of the method at the present time: to determine the component abilities entering into a "general ability," and to determine the diagnostic value of different tests. Kelley, Rosenow, Wendle and Wyman, Higbie, Toops, and Gray are among the chief users of the method. But it is to Kelley that we owe the real impetus for the movement (and to Thorndike for his insight in pointing the course of development), both in making the pioneer use of the method (Ref. 29) and in developing the tables by which the labor and time of computation can be so materially shortened. Rosenow has thrown helpful light on our thinking about scientific methods and he, too, has contributed important time-and-labor-saving suggestions (Ref. 5).

Just what can we do with partial correlation? What is the significance of the term "partial?" Let us take a common-place example, say Rosenow's illustration of finding the relation between yield of crops (called x_1), rainfall (x_2) and sunshine (x_3). The coefficient of correlation between yield and rainfall alone would

be complicated by the unaccounted-for factor of sunshine. So we desire to "eliminate," or "hold constant," the effect of sunshine. We do this by finding the combined effect of the sunshine and rainfall on yield by adding the yield due to rain with sunshine constant to the yield due to sunshine with rain constant. As an equation, it reads:

$$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$$

in which x_1 is the yield, x_2 the rainfall, and x_3 the sunshine. To understand $b_{12.3}$ and $b_{13.2}$, recall that the equation for correlation between the variables x and y is

$$y = b_1 x, \quad \text{or} \quad y = r\frac{\sigma_y}{\sigma_x} x$$

where $b_1 = r\frac{\sigma_y}{\sigma_x}$, and is called the *regression coefficient.* Now, since a third variable is added, we need a scheme of notation. The correlation between yield and rain we will call r_{12}; the correlation between yield and sunshine r_{13}; the correlation between rain and sunshine r_{23}. These subscripts enable us to tell which variables are being related and which ones are held constant, *i.e.*, the effects of which ones are eliminated. A coefficient of "partial" correlation will have the notation: $r_{12.345\ldots.n}$. The subscripts to the left of the point (12) are *primary* and denote the variables which are being correlated; those to the left are *secondary* and are "eliminated" variables. x_1 is called the *dependent variable,* x_2 and x_3 the *independent variables.*

In the complete equation (Ref. 29):

$$\sigma_{1.23} = \sigma_1 \sqrt{1 - r^2_{13}} \quad \sqrt{1 - r^2_{12.3}}$$

$$\sigma_{2.13} = \sigma_2 \sqrt{1 - r^2_{23}} \sqrt{1 - r^2_{12.3}}$$

$$\sigma_{3.12} = \sigma_3 \sqrt{1 - r^2_{23}} \sqrt{1 - r^2_{13.2}}$$

This shows that to find the relative extent of the influence of each variable (shown by b_1 and b_2) it is necessary to compute all the "coefficients of zero order," *e.g.*, r_{12}, r_{13}, r_{23} and the coefficients of the first order $r_{12.3}$, $r_{13.2}$, etc.

What we do in multiple correlation, therefore, is to determine the correlation that exists between actual values of x_1 and values estimated from the equation of partial regression:

$$x_1 = b_{12.3}\, x_2 + b_{13.2}\, x_3$$

Just as with two variables, so with three or *n* variables we obtain a coefficient of multiple correlation, R, which is a measure of the closeness with which we can estimate x_1 from $x_2, x_3, x_4, \ldots x_n$.

Limitation of space prohibits presenting the details of computation. Suffice it to say that Kelley (Ref. 29) and Rosenow (Ref. 5) have developed short methods and tables by which computation is extraordinarily facilitated. The advanced student should master the methods as set forth in these two treatments.

2. Limitations of Multiple Correlation Methods

The most serious limitation that the worker who uses partial regression equations should have in mind is that it assumes that the influence of the independent variables x_2 and x_3 on the dependent variable x_1 is additive. Probably this seldom actually obtains. Thurstone's homely illustration (Ref. 12) of the relation between the volume (v) of a box and the length (l), width (w), and the depth (d) makes the point clear. The true relation is given by

$$v = k.d.w.l,$$

but the best expression we could obtain by multiple correlation would be of the form

$$v = k_1 d + k_2 d + k_3 l.$$

We have no known methods of handling a situation of this kind. Furthermore, we know nothing of the manner of combination of the constituents of gross mental functions.

The second limitation is that partial correlation is based on the assumption of linear relationship. For any non-linear relationship (and it may be that they will be found for mental functions) such an assumption leads to a coefficient and an equation which are totally fictitious measures of the true correlation. It is possible, however, to rectify a non-linear regression by mathematical devices used with empirical equations (see Thurstone, 12).

3. Empirical Equations as Predictive Measures

The correspondence of two series of values can be expressed in three ways: (1) as a table of correlated values; (2) as a line of most probable relationship from a scatter diagram of observed measures; (3) as the equation of such a line of relationship. Education is now using all three of these methods, the last one only recently. The regression equation already mentioned is an illustration of our progress in the statistical treatment of such data. There are three methods by which the observed data of a correlation table may be expressed as an equation: (1) The simplest method is to fit a line by inspection to the points of the table, measuring the y–intercept and the slope of the line and obtaining an equation of the form $y = mx + b$; (2) the second is the method of the regression equation (see Ref. 26); (3) the third is the method of least squares which gives the values of the constants a and b in the equation $y = a + bx$, and from which we can predict the most probable value of y from a known value of x (see Thurstone, 12).

A new path of development has been blazed out by Thurstone's pioneer attempt to describe the course of the learning process by fitting empirical equations to the data of learning (Ref. 12). Thorndike suggested years ago the feasibility of determining the equations of basic learning curves and called attention to the fundamental form of those so far reported (Ref. 27). Thurstone, after trying about 40 different equations on published learning curves, selected a hyperbola of the form

$$Y = \frac{L\,(X + P)}{(X + P) + R}$$

in which $Y =$ attainment, $X =$ formal practice, $P =$ equivalent previous practice, $L =$ limit of practice, and $R =$ rate of learning. He illustrates how such a curve can be rectified by turning the equation into the form $X + (R + P) = \dfrac{L\,(X + P)}{Y}$, which is linear, if values of $\dfrac{(X + P)}{Y}$ are plotted against values of X. If a curve be so rectified, the constants L, R, and P can be determined by any one of several methods, four of which he illustrates.

Here, then, is another illustration of the way in which the science of education is refining the statistical treatment of its data and perfecting its method of describing observed facts and of determining its basic laws.

SECTION III.—ANNOTATED BIBLIOGRAPHY OF RECENT DEVELOPMENTS IN THE USE OF STATICTICAL METHODS IN EDUCATION[7]

A. Statistical Methods Employed in Determining Reliability of Tests

1. Kelley, T. L. "The reliability of test scores." *Jour. of Educ. Research*, May, 1921, 370-379.

   An important summary of possible methods of determining reliability with evaluation of each method. Emphasizes importance of probable errors of estimates.

2. Otis, Arthur S., and Knollin, H. E. "Reliability of Binet Scale and pedagogical scales." *Jour. of Educ. Research*, September, 1921, 121-142.

   Largely a discussion of the value and technique of using probable errors of scores to measure reliability of tests. Compares this method with improper uses of coefficients of correlation, and shows influence of greater variability of some school groups on results obtained. Reports the use and derivations of (1) a difference formula for correlation, (2) a formula for the probable error of a single measure in terms of median difference between measures, (3) a formula for the probable error of half a scale.

3. Kelley T. L. "The measurement of overlapping" *Jour. of Educ. Psych.*, November, 1919, 458-461.

   Points out incorrectness of all measures of overlapping reported to 1919, and need for using formula for standard deviation of an infinitely large number of similar tests when the standard deviation and the coefficient of reliability of the single tests is known.

---

[7] The National Research Council has in preparation a volume that will bring together in a condensed form practically everything that has been done on the application of statistics in the various fields of research. This handbook, with its comprehensive bibliography, may be expected to appear some time in 1922.—*Editor.*

B. DETAILED DEVELOPMENT (WITHOUT THE CALCULUS) OF THE THEORY OF MULTIPLE CORRELATION

4. Yule, G. A. *An Introduction to the Theory of Statistics,* pp. 229-253.

C. APPLICATION TO EDUCATION AND EDUCATIONAL PSYCHOLOGY OF THE THEORY OF MULTIPLE AND PARTIAL CORRELATION

5. Rosenow, Curt. "The analysis of mental functions." *Psych. Monographs;* Vol. XXIV, No. 5, 1917.

   Contains excellent exhibit of possible uses of partial correlation in the analysis of mental abilities and a non-mathematical evaluation of the theory itself. This is a pioneer application of partial correlation in this field and should be read by all students of that statistical method. Appendix contains directions for computation of coefficients by short methods which make possible very large reductions in labor and time.

6. Kelley, T. L. *Tables: To Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations.* Bulletin of the University of Texas, 1916, No. 27, Austin, Texas.

   A technical statement of what partial coefficients and regression equations are, and how they can be used, with outlines and illustrations of the procedure to be followed in calculating them. By means of Kelley's tables a reduction of about 80 percent is effected in the labor of computation. The student should know both Kelley's and Rosenow's (No. 5) methods.

7. Kelley, T. L. *Educational Guidance.* Teachers College, Columbia University, Contributions to Education, No. 71, 1914.

   The pioneer use of partial correlation in the analysis and prediction of ability of high-school pupils. Kelley is the first American educational psychologist to utilize methods of multiple correlation. A very technical statistical discussion.

8. Higbie, E. C. *An Objective Method for Determining Certain Fundamental Principles in Secondary Agricultural Education.* (Privately published, doctorate dissertation, Teachers College, Columbia University.)

   Uses partial correlation to determine the contribution of different traits (*e.g.*, native intelligence, managerial ability, mechanical ability,

physical ability, and others) to success in farming when financial success and community value are regarded as two criteria.

9. Toops, H. A. *Trade Tests in Education.* Teachers College, Columbia University, Contributions to Education, No. 115, 1921.

   Employs partial correlation to determine relative value of tests for ability in English, arithmetic, filing, use of switchboard, stenography and typewriting, general adaptability, personality, appearance, etc., in predicting trade abilities. Uses formulas for reliability of tests. Gives technical summary of statistical methods of correlation.

10. Gray, C. T. *A Score Card for the Measurement of Handwriting.* Bulletin No. 37, 1915, of the University of Texas, Austin, Texas.

    Employs multiple correlation to determine weights that should be given to nine contributory elements of handwriting. An early use of partial correlation, stimulated by Kelley.

D. Important Illustrations of the Practical Use of Multiple Correlation in Predicting Future Conditions

11. Moore, H. L. *Forecasting the Yield and the Price of Cotton.* MacMillan, 1917, New York.

    A pioneer use of correlation in economic prediction. Shows that it is possible to employ multiple correlation and regression equations with three variables to forecast the yield of cotton more accurately from the data of rainfall and temperature than is done by the elaborate official machinery now employed by the Federal Department of Agriculture. Presents a good brief resume of the mathematics of correlation. Has important values for the student of educational and psychological statistics.

E. The Use of Curve-Fitting as a Means of Prediction

12. Thurstone, L. L. "The learning curve equation." *Psych. Monographs,* Vol. XXVI, No. 3, 1919.

    The pioneer investigation of curve fitting in educational psychology. Primarily an illustration of how to fit empirical equations to learning data to determine exact laws of prediction. Refers to partial correlation methods in introduction.

F. New Formulas for Correlation

13. Kelley, T. L. "A simplified method of using scaled data for purposes of testing." *School and Society*, July 1, 1916, 34-37; July 8, 71-74.

Reports formula for correlation between score in one test and the estimated average score in a succession of tests.

14. Otis, Arthur S. "The reliability of spelling scales involving a 'deviation formula' for correlation." *School and Society*, 1916, Oct. 28, pp. 677-683; Nov. 4, pp. 716-722; Nov. 11, pp. 750-760.

Reports an elaborate statistical analysis of spelling scales and a new coefficient of correlation based upon a "curve of rank relation."

15. Ruml, B. "The reliability of mental tests in the division of an academic group." *Psych. Monographs*, Vol. XXIV, No. 4, 1917.

Reports statistical methods of using mental tests for classifying pupils; use of Pearson's "Scale of Intelligence." Of interest to student of statistics because it reports a *rank-tangential coefficient* (t) for the relation between a continuous variable and a variable divided at some point into alternative categories.

16. Ruml, B. "The measurement of the efficiency of mental tests." *Psych. Rev.*, November, 1916, 501-507.

Formula for determining practical efficiency of a test.

G. The Use of Brown's Formula

17. Brown, Wm. *The Essentials of Mental Measurement.* Cambridge University Press, London, England, 1911 (pp. 101-102).

Gives derivation and use of the formula.

18. Burgess, May Ayres. *The Measurement of Silent Reading.*
4 Russell Sage Foundation, New York City, 1921 (pp. 128-132).

5 Non-technical discussion of the formula and what its use really implies. Valuable.

19. Gates, Arthur I. "An experimental and statistical study of reading and reading tests." *Jour. Educ. Psych.*, September, October, and November, 1921.

An elaborate study of inter-correlations between different tests of "reading ability," and use of Brown's formula for determining reliability.

20. Wyman, J. B., and Wendle, Miriam. "What is reading ability?" *Jour. Educ. Psych.*, December, 1921, 518-531.

Elaborate use of partial correlation and reliability formulae for tests of elements entering into reading ability. Reports first use of Kelley's formula for the probable error of a coefficient of correlation corrected for attenuation, together with criticism of Spearman's "corrected coefficients."

H. Short Statistical Methods

1. Computation of Product-Moment Coefficients of Correlation

21. Ayres, L. P. "A shorter method for computing the coefficient of correlation." *Jour. Educ. Research,* March, 1920, 216-21.

Helpful only when large numbers of coefficients are to be computed and statistical machines are to be used.

22. Ayres, L. P. "The application of tables of distribution of a shorter method of computing coefficients of correlation." *Jour. Educ. Research,* April, 1920, 295-298.

23. Ayres, L. P. "Substituting small numbers for large ones in the computation of coefficients of correlation." *Jour. Educ. Research,* June, 1920, 502-504.

24. Buckingham, B. R. "Proof of Dr. Ayres' Formula." (Editorial). *Jour. Educ. Research,* June, 1920, 505-507.

25. Ayres, L. P. "The correlation ratio." *Jour. Educ. Research,* June, 1920, 452-457.

A short method of computing the correlation ratio, n.

26. Rugg, H. O. *Statistical Methods Applied to Education.* Houghton-Mifflin Company, 1917.

27. Thorndike, E. L. *An Introduction to the Theory of Mental and Social Measurements.* Teachers College, Columbia University, 1913.

2. Computation of Rank-Difference Coefficients of Correlation

28. The Scott Company Laboratory, Philadelphia. "Tables to facilitate the computation of coefficients of correlation by the rank-difference method." *Jour. Applied Psych.,* June-September, 1920, 115-125.

3. Computation of Partial Coefficients of Correlation and Regression Equations

29. Kelley, T. L. *Tables to Facilitate the Calculation of Partial Coefficients of Correlation and Regression Equations.* Bulletin No. 27, 1916, University of Texas, Austin, Texas.

TESTS FOR THE KINDERGARTEN AND GRADES I TO III

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
An Absolute Intelligence Scale (Experimental Edition)	Grace Arthur and Herbert Woodrow	1. Immediate Memory 2. Easy opposites 3. Hard opposites 4. Substitution 5. Word building 6. Language completion 7. Anagrams 8. Cancellation 9. Comprehension	Ages 6.5 to 13.5 years				Grace Arthur and Herbert Woodrow, "An Absolute Intelligence Scale. A study in method." *Jour. Applied Psych.*, 3: June, 1919, 118-137.
Cole and Vincent Group Intelligence Test for School Entrance	Cole and Vincent		Kindergarten and first grade Ages 5 and 6		Bureau of Educ. Meas. and Stand. Kansas State Normal Sch. Emporia, Kan.	25 copies with stencils and directions, $1.50	
The Dearborn Group Tests of Intelligence, Series I. General Examinations 1, 2 and 3	Walter F. Dearborn	Forty-two pictorial tests, many adapted for group work from Binet	Grades 1-3	Three periods of 25 min. each	J. B. Lippincott Co., Phila., Pa.	25 tests, $1.70 100 tests, $6.00 Instructions booklet, 25c Stencils separately	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Detroit First-Grade Intelligence Test, Group Test A	Anna M. Engel	Ten pictorial tests: 1. Acquired information 2. Appreciation of similarities 3. Memory 4. Appreciation of absurdities 5. Comparisons 6. Relationships 7. Sympathy 8. Designs 9. Ability to count 10. Simple directions regarding spatial relations	First-grade entrants	20-30 min.	World Book Co., Yonkers, N. Y.	25 Booklets with record sheets, $1.50 Examiners' Guide, 10c Specimen set, 15c	"Use of a First-Grade Intelligence Test in Detroit," *Jour. Educ. Research*, 3: Jan., 1921, 76-77. Also *Detroit Educ. Bulletin*, Nov., 1920.
Intelligence Examination Delta 1 (Used in the Virginia School Survey)	M. E. Haggerty	Six fore-exercises and six tests (5 non-verbal) 1. Directions 2. Copying designs 3. Picture completion 4. Picture comparison 5. Symbol-digit 6. Word comparison	Grades 1-3	30 min.	World Book Co., Yonkers, N. Y.	25 booklets with record sheet, $1.50 Key 15c Manual Directions, 40c Specimen sot, 65c	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Holley Picture Completion Test for Primary Grades	C. E. Holley	Twenty pictures to complete, preceded by a fore-exercise. One form	Grades 1-3	15 min.	The Public School Pub. Co., Bloomington, Ill.	$1.25 per 100 Sample set, 10c	
Kingsbury Primary Group Intelligence Scale, Form "A"	Forrest A. Kingsbury	Four tests 1. Right answers 2. Opposites 3. Completion 4. Form	Grades 1-4	25 min.	The Public School Pub. Co., Bloomington, Illinois	$2.50 per 100 Single copy, 10c	
Group Intelligence Scale for Primary Grades	Frances Lowell	Twenty-five tests, taken from Binet, Kuhlmann, *et al*, arranged five for each year, 5 to 9	Grades 1-3 or Ages 5-9, inclusive				Frances Lowell. "A group intelligence scale for primary grades." *Jour. Applied Psych.* 3: Sept., 1919, 215-247
Primary Group Test of General Ability (Gunnison Primary Test-A)	H. T. Manuel	Pictorial (Experimental Edition)	Grades 1-3, inclusive	40-45 minutes on each of two successive days	Dr. H. T. Manuel, Director of Educational Research, Colorado State Normal, Gunnison, Colo.	$5.00 per 100 Specimen set, 60c Manual of Directions, 50c	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Myers' Mental Measure	Caroline E. and Garry C. Myers	Four pictorial tests 1. Directions 2. Picture completion 3. Learning 4. Common elements	For all grades from the kindergarten to the university	20 min.	Newson & Co. 73 5th Ave. New York, or 623 So. Wabash Ave., Chicago, Ill.	1 copy 10c 12 copies \$1.00 100 copies \$5.00 1000 copies \$40.00 Measuring Minds—An Examiner's Manual, 1 copy 80c 12 copies \$7.20 100 copies \$54.00	Caroline E. and Garry C. Myers. *Measuring Minds*, Newson and Co., New York. C. E. and G. C. Myers, "A group intelligence test," *School and Society*, 10: Sept. 20, 1919, 355-360
Otis Group Intelligence Scale Primary Examination	Arthur S. Otis	Eight tests (six non-verbal) 1. Directions 2. Associations 3. Picture completion 4. Maze 5. Picture sequence 6. Similarities 7. Synonym-antonym 8. Common sense Two forms available	Grades 1-4	25 min.	World Book Co., Yonkers, N. Y.	25 booklets, with record sheet, \$1.50 Manual of Directions, 40c Specimen set, 75c, postpaid	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Mental Survey Tests; Primer Scale	Mrs. Sidney D. Pressey	Four pictorial tests 1. Dot pattern 2. Classification 3. Picture test form-board 4. Picture absurdities	Grades 1-3	25 min.	Public Sch. Publishing Co., Bloomington, Ill.	$2.50 per 100 Sample set, 10c	Luella W. Pressey, "A group scale of intelligence for use in the first three grades." *Jour. Educ. Research*, 1: Apr., 1920, 285-294
The Mentimeters	M. R. Trabue and F. P. Stockbridge	School Group 2a (Mentimeters 2, 3, 5, 16, 20, 23, 28, 18) comprises 1. Pictorial absurdities 2. Maze threading 3. Geometric figures 4. Opposites 5. Reading directions 6. Completion 7. Arithmetic 8. Range of information Other mentimeters described in book	Applicable to all persons from infants to university graduates	Varies with tests selected	Doubleday, Page & Co., Garden City, L. I.	Fifty pamphlets with stencils, record sheet, and quotient tables, $3.25, plus transportation Half sets for 25 pupils, $1.75, plus transportation Specimen set, 25c The Mentimeter Quotients supplied separately, free to the purchaser	M. R. Trabue and F. P. Stockbridge *Measure your Mind: The Mentimeter and How to Use It.* Doubleday, Page & Co., 1920, Garden City, N. Y.

II. Tests Primarily for the Elementary Grades

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
The Dearborn Group Tests of Intelligence, Series II, General Examinations 4 and 5	Walter F. Dearborn	Ten tests (7 non-verbal) 1. Picture sequences 2. Word sequences 3. Form completion 4. Opposite completion 5. Memory ladder 6. Picture symbols 7. Mazes 8. Disarranged proverbs 9. Faulty pictures 10. Number problems	Grades 4-9	Examinations 4 and 5 will take 50 min. each. Should be given to each grade with an interval of time (preferably a class period) between them	J. B. Lippincott & Co., Phila., Pa.	25 booklets $1.70; 100 booklets $6; 1000 booklets $58. Instruction booklet, 25c.	
Haggerty's Intelligence Examination, Delta 2 (Used in Virginia School Survey)	M. E. Haggerty	1. Sentence reading 2. Arithmetic 3. Picture completion 4. Synonym-antonym 5. Common-sense 6. General information	Grades 3-9 Ages 8-15	35 min.	World Book Co., Yonkers, N. Y.	25 booklets, with record sheet, $1.50 Specimen set, 65c Scoring key, 10c Manual of Directions, 40c	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
Holley's Sentence Vocabulary Scale, Series 3B	C. E. Holley	See High-School Tests					
Myers' Mental Measure		See Primary Tests					
Illinois Examination, I	W. S. Monroe and B. R. Buckingham	A group intelligence scale followed by 8 arithmetic tests and the Monroe Reading Test. The intelligence tests are 1. Analogies 2. Arithmetic problems 3. Sentence vocabulary (information) 4. Substitution 5. Verbal ingenuity (Disarranged sentences) 6. Arithmetical ingenuity 7. Synonym-antonym	Grades 3, 4, 5.	50-60 min. (27 min. actual working time)	Public Sch. Pub. Co., Bloomington, Ill.	Sample set including Hand Book 25c Illinois Examination II, $4.00 per 100 Intelligence test also printed separately, $2.00 per 100	B. R. Buckingham and W. S. Monroe, "A testing program for elementary schools," *Jour. Educ. Research*, 2: Sept., 1920, 521-532

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
Illinois Examination II	W. S. Monroe and B. R. Buckingham	Intelligence examination same as Illinois Examination I, Arithmetic scale has 7 tests. Remainder of pamphlet has 7 tests of arithmetic and tests of reading. Two forms available.	Grades 6, 7, 8.	60 min. (37½ min. actual working time)	Public Sch. Pub. Co., Bloomington, Ill.		
National Intelligence Tests, Scale A-Scale B	Prepared by M. E. Haggerty L. M. Terman E. L. Thorndike G. M. Whipple R. M. Yerkes	Two scales of 5 tests each. Scale A: 1. Arithmetic problems 2. Sentence completion 3. Logical selection 4. Synonym-antonym 5. Symbol-digit Scale B: 1. Computation 2. Information 3. Vocabulary 4. Analogies 5. Comparison (Two forms of each scale available—three more in preparation)	Grades 3-8.	30-35 min.	World Book Co., Yonkers, N. Y.	Either scale prices same 25 booklets, with keys and record sheet, $1.60 Manual of Directions, 25c Specimen Set, 50c, postpaid	(1) G. M. Whipple, "The National Intelligence Tests," *Jour. Educ. Research*, 4: 1921, 16-31 (2) L. M. Terman and Edith D. Whitmire, "Age and grade norms for the National Intelligence Tests," Scales A and B, *Jour. Educ. Research*, 3: Feb. 1921, 124-132

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
Otis Group Intelligence Scale, Advanced Examination	Arthur S. Otis	1. Following directions 2. Opposites 3. Disarranged sentences 4. Proverbs 5. Arithmetic 6. Geometric figures 7. Analogies 8. Similarities 9. Narrative completion 10. Memory Two forms available	Grades 5-12	60 min.	World Book Co., Yonkers, N. Y.	25 booklets, $1.50 Examiner's key, 25c Manual of directions, 40c Specimen set, 85c	(1) A. S. Otis, "An absolute point scale for the group measurement of intelligence," *Jour. Educ. Psych.* 9: May-June, 1918, 239-61, 323-48. (2) W. D. Armentrout, "Classification of junior high school pupils by the Otis Scale," *Jour. Educ. Psych.* March, 1920, 165-168. (3) V. E. Dickson and John J. Norton, "The Otis Group Intelligence Scale applied to the elementary school graduating classes of Oakland, Cal." *Jour. Educ. Research,* 3: Feb. 1921, 106-115. (4) S. S. Colvin, "Recent results obtained from the Otis Group Intelligence Scale," *Jour. Educ. Research,* 3: Jan. 1921, 1-12.

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
Pintners' The Mental Survey Tests	Rudolph Pintner	Six tests taken largely from Whipple's *Manual of Mental and Physical Tests* and standardized on 3000 children: 1. Rote memory 2. Digit-symbol 3. Symbol-digit 4. Word building 5. Opposites 6. Cancellation	Ages 6-16		C. H. Stoelting Co., 3037 Carroll Ave. Chicago, Ill.		R. Pintner, *The Mental Survey,* D. Appleton & Co., New York
Non-Language Mental Tests	Rudolph Pintner	Six tests designed to be used with an accompanying set of educational tests: 1. Imitation 2. Easy learning 3. Hard learning 4. Drawing completion 5. Reversed drawings 6. Picture construction	Best results in Grades 3-8. Can be used in second grade and in the high school, but not so reliably	40-45 min.	College Book Co., Columbus, O.	Per 100, $4.00 Manual of Directions, 50c Sample set, including manual, stencils, and 5 copies of both mental and educational blanks, $1.40	(1) R. Pintner and Helen Marshall, "A combined mental-educational survey," *Jour. Educ. Psych.*, 12: Jan. 1921, 32-43 and Feb., 82-91. (2) R. Pintner, "A non-language group intelligence test," *Jour. Applied Psych.*, 3: Sept. 1919, 199-214

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Price	References
Mental Survey Scales, "Cross-out" Tests, Schedule E	Sidney L. Pressey	1. Verbal ingenuity 2. Logical judgment 3. Arith. ingenuity 4. Moral judgment	Grades 4-12 Ages 10-17 Norms for Low 4 to High 8		Univ. of Indiana, Bloomington, Ind.	100, with directions, score sheet, record blank, and norms, $1.00 plus postage	(1) S. L. and L. W. Pressey, "'Cross-out' test, with suggestions as to a group scale of the emotions." *Jour. Applied Psych.*, 3: June, 1919, 2: 138-150 (2) S. L. Pressey "A brief group scale of intelligence for use in school surveys," *Jour. Educ. Psych.*, 11: Feb. 1920, 89-100.
A Group Point Scale for Measuring General Intelligence	S. L. and L. W. Pressey	1. Rote Memory 2. Logical selection 3. Arithmetic 4. Opposites 5. Logical memory 6. Word completion 7. Moral classification 8. Dissected sentences 9. Practical information 10. Analogies	Ages 8-17				S. L. and L. W. Pressey, "A group point scale for measuring general intelligence, with first results from 1100 school children," *Jour. Applied Psych.* 2: Sept., 1918, 250-269.

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Omaha Group Test of Intelligence	Paul B. Stevenson	Eight tests (from 5 to 15 exercises each) 1. Likeness and difference 2. Correction of statement 3. Arith. problems 4. Disarranged sentences 5. Following directions 6. Synonym-antonym 7. Analogies 8. Range of information	Grades 3-9	20 min.	Bureau of Research, Univ. of Omaha, Omaha, Neb.		
N. Y. Board of Charities, 24 Standardized Mental Tests	Based on experiments conducted by Dr. Gertrude E. Hall	Twenty-four tests			C. H. Stoelting Co., 3037 Carroll Ave., Chicago, Ill.		Eugenics and Social Welfare Bulletins, Nos. V and VIII
A Group Classification Test	W. W. Theisen and Cecile White Flemming	1. Following directions 2. Synonym-antonym 3. Arithmetic 4. Common sense 5. Completion 6. Analogies 7. Number completion 8. Information	Grades 5-12	15 min.	Bureau of Publications, Teachers College, Columbia Univ., N. Y.		W. W. Theisen, "Group intelligence tests," *Jour. Educ. Research*,

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
A Standardized Group Examination of Intelligence Independent of Language	E. L. Thorndike	Eight tests derived largely from Army Beta 1. Digit-symbol 2. Lines dividing surfaces 3. Picture completion 4. Form completion 5. Picture analogies 6. Spatial relations 7. Easy computation	Applied to groups ranging from mental age 7½ to superior adults	60 min. (42 min. working time)	Bureau of Publications, Teachers College, Columbia Univ., New York		E. L. Thorndike, "A standardized group examination of intelligence independent of language," *Jour. Applied Psych.* 3: Mar., 1919, 13-32
Mentimeters	Trabue and Stockbridge	See Primary Tests					
Whipple's Group Tests for Grammar Grades (Revised Edition, 1920)	G. M. Whipple	1. Computation 2. Completion 3. Comparison 4. Logical selection 5. Punched holes 6. Equivalent proverbs	Grades 4-8	35-40 min.	C. H. Stoelting Co., Chicago, and Public Sch. Pub. Co., Bloomington, Ill. Also by the compiler	25 for $1.75 Single copy, 10c Manual, 20c	(1) G. M. Whipple, *Classes for Gifted Children*, Public Sch. Pub. Co., Bloomington, Ill. (2) Helen Davis, "Validity of the Whipple Group Tests for Grammar Grades," to appear in *Jour. Educ. Research*

III. TESTS PRIMARILY FOR THE HIGH SCHOOL AND COLLEGE

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Army Alpha	Committee of American Psych. Ass'n; W. V. Bingham H. H. Goddard T. H. Haines L. M. Terman Lyman Wells G. M. Whipple R. M. Yerkes	1. Following directions 2. Arithmetic problems 3. Practical judgment 4. Synonym-antonym 5. Disarranged sentences 6. Number series completion 7. Analogies 8. Information Originally used in five forms	Secondary schools and colleges and adults generally	40-50 min.	Originally Division of Psychology, Surgeon General's Office War Dept., Washington, D. C. At present Bureau of Educ. Measurements and Standards State Normal Sch., Emporia, Kansas (5 forms). Also, Bureau of Educ. Reference and Research, Univ. of Michigan, (Forms 6 and 9 only)	100 booklets, $3.00; Manual of Instructions each 75c Sample set, 80c	(1) C. S. Yoakum, and R. M. Yerkes, *Army Mental Tests*, Henry Holt & Co. 1920. (2) *Psychological Examining in the United States Army*. Edited by R. M. Yerkes, Memoirs of the Nat'l Academy of Sciences, 5: 1921, 890 pp. (3) I. N. Madsen and R. H. Sylvester. "High-school students' intelligence rating according to the Army Alpha test." *School and Society*, 10: Oct. 4, 1919, 407-410. (4) I. N. Madsen, "High-school students' intelligence ratings according to the Army Alpha test." *Sch. and Soc.*, 11: March 6, 1920, 298-300.

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Army Alpha (continued)							(5) E. L. Noble, "University students' intelligence ratings according to the Army Alpha test." *Sch. and Soc.*, 11: Feb. 21, 1920, 233-237. (6) M. J. Van Wagenen. "Our schools as measused by the army tests." *Educ. Adm. and Superv.* 5: April, 1919, 63-76. (7) H. H. Goddard, *Human Efficiency and Levels of Intelligence.* Princeton University Press, 1921
"Test VI" (Sometimes referred to as "Scrambled Alpha") Two forms available	Bureau of Personnel Research	184 items from Army Alpha, arranged in 'spiral' form: includes 1. Arithmetic 2. Synonym-antonym 3. Disarranged sentences 4. Number series completion 5. Analogies 6. Information	Secondary schools and colleges and adults generally	25 min. working time	Bureau of Personnel Research, Carnegie Institute of Technology, Pittsburgh, Pa.	Limited number have been sold at 4 cents per copy	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Brown University Psychological Examination	Stephen S. Colvin	Fore-Exercise A 1. Completion 2. Definitions 3. Opposites 4. Analogies Exercise B 1. Completion 2. Definitions 3. Opposites 4. Analogies 5. Arithmetic Exercise C Same five tests with different items	College Freshmen	Exercise A, 10 min. Exercise B, 30 min. Exercise C, 30 min.	S. S. Colvin, School of Education, Brown Univ., Providence, R. I.	Not yet on the market, but tests have been sold at a cost of 10c per set. Sample set, 10c	(1) S. S. Colvin, "Psychological tests at Brown University," *Sch. and Society*, 10: July 5, 1919, 27-30 (2) S. S. Colvin, "Validity of psychological tests for college entrance," *Educ. Rev.*, 60: June, 1920, 7-17 (3) S. S. Colvin, "Purposes and methods of psychological tests in colleges," *Educ.* 40: March, 1920, 404-416

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Junior and Senior High-School Classification Test	J. Crosby Chapman and J. B. Welles	1. Opposites 2. Arith. problems 3. Information	Junior and Senior high school	11 min.	The Dobson-Evans Co., 305-7 No. Front St., Columbus, O.	100 sheets, including directions, etc., $3.00 1000 sheets, $25.00	
Holley's Sentence Vocabulary Scale, Series 3 B	C. E. Holley	Based on Stanford-Binet vocabulary, but said to be a fair measure of intelligence	Grades 7-12		Public Sch. Pub. Co., Bloomington, Ill.	80c per 100, Sample set 6c	*First Annual Report of Bureau of Educ. Research*, 1918-19, Univ. of Illinois
Miller Mental Ability Test	W. S. Miller	1. Disarranged sentences combined with directions test 2. Controlled association (Vocabulary) 3. Analogies One form available	Grades 7-12	30 min.	World Book Co., Yonkers, N. Y.	25 blanks with key, $1.00 Percentile graphs, 10c Age-grade score sheet, 10c Manual of Directions 20c Specimen set, 30c	W. S. Miller, "The administrative use of intelligence tests in the high school." See elsewhere in this *Yearbook*
Myers' Mental Measure		See Primary Tests					
Otis Group Intelligence Scale, Higher Examination, Forms A and B	Arthur S. Otis	Omnibus type, self-administering	High School and college	40 min.	World Book Co., Yonkers, N. Y.	25 booklets, with directions, etc., $1.00 Specimen set, 85c, postpaid Manual of Directions, 40c	

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Roback Mentality Tests for Superior Adults (prepared by the author for Simmons College)	Dr. A. A. Roback, Emerson Hall, Harvard Univ.	1. Abstraction 2. Problems 3. Analogy 4. Relations 5. Insertion 6. Reference 7. Opposites 8. Acumen 9. Subsumption 10. Directions (in later edition interpretation test substituted) 11. Judgment 12 Cryptogram	College students	185 min.	J. C. Miller, Jr., 5 Lauriat Place, Medford, Mass.	1 copy 50c 3 copies $1.00 12 copies $4.00 50 copies $15.00 100 copies $25.00 500 copies $100.00 Booklet with full scoring instructions, 35c	
Group Tests of Intelligence	David Camp Rogers	1. Logical conclusions 2. Delayed recall of ideas 3. Information 4. Arith. problems 5. Immediate recall of ideas 6. Substitution 7. Similar relation 8. Completion 9. Absurdities 10. Following directions 11. Train of associations	College freshmen	60 min. (44 min. working time)	Specimen copies in limited quantities can be obtained without charge by addressing D. C. Rogers, 319 Elm St., Northampton, Mass.	Not recommended for general use, and no copies for sale available. Devised for experimental purposes	D. C. Rogers, "Intelligence examinations and college entrance," *The Smith Alumnae Quarterly*, 13: Nov., 1921

Title	Compiled by	Number and Nature of Tests	Range—Ages or Grades	Time to Apply	Publisher	Prices	References
Terman's Group Test of Mental Ability	Lewis M. Terman	1. Information 2. Best answer 3. Word meaning 4. Logical selection 5. Arithmetic 6. Sentence meaning 7. Analogies 8. Mixed 9. Classification 10. Number series	Grades 7-12	35 min.	World Book Co., Yonkers, N. Y.	25 booklets including 1 Manual and 1 Key, $1.60 Specimen set 15c	
Thorndike Intelligence Examinations (Sometimes listed as "Thorndike Intelligence *Examinations* for College Entrance," sometimes as, "Thorndike Intelligence Examination for High-School Graduates")	E. L. Thorndike	Complete set comprises five booklets Part I, Practice Form, contains 13 Tests: 1. Directions 2. Disarranged sentences 3. Computation 4. Arith. problems 5. Information 6. Synonym-antonym 7. Practical judgment 8. Number series 9. Analogies 10. Number discrimination 11. Absurdities	High-school graduates and college students	2 hrs. and 40 min. working time plus 30 min. additional for administration	Bureau of Publications, Teachers College, N. Y.	Sample set, 50c postpaid Current issues, 50 sets, $50.00 Back issues, 25 sets, $20.00 Three issues yearly	(1) E. L. Thorndike, "Intelligence examinations for college entrance," *Jour. Educ. Research*, 1: May, 1920, 329-37 (2) E. L. Thorndike, "Tests of intelligence," *Sch. and Society*, 9: Feb., 1919, 189-195

Title	Compiled by	Number and Nature of Tests	Range of Ages or Grades	Time to Apply	Publisher	Prices	References
Thorndike Intelligence Examinations (continued)		12. Logical conclusions 13. Recognitory memory Two other test booklets contain same 13 tests. Part II contains eight tests: 1. Reading 2. Sentence completion 3. Picture completion 4. Picture analogies 5. Form proportion 6. Algebra 7. Practical information 8. Information Part III has two reading tests and another sentence completion					
Psychological Examination for College Freshmen and High-School Seniors (Test IV in series of six tests for engineering freshmen)	L. L. Thurstone	168 items, including 1. Information 2. Analogies 3. Sentence completion 4. Syllogisms 5. Quotations 6. Number completion	Secondary schools and colleges	30 min. working time	C. H. Stoelting Co., 3037 Carroll Ave., Chicago, Ill.	$15.00 per 100 copies	(1) L. L. Thurstone, "Mental tests for college entrance," *Jour. Educ. Psych.*, 10: 1919, 129-142 (2) "A cycle-omnibus test for college students." *Jour. Educ. Research*, 4: Nov. 1921, 265-278
Mentimeters	Trabue and Stockbridge	See Primary Tests					

CHAPTER I

INTELLIGENCE TESTS AND INDIVIDUAL PROGRESS IN SCHOOL WORK

Henry W. Holmes
Dean of the Graduate School of Education, Harvard University, Cambridge, Massachusetts

Every new movement in education calls for someone to repeat the warning dictum of Emerson: An expense of ends to means is fate. We are too often satisfied to exemplify a method or use a means without critical examination of the ends we are serving; and whenever our zeal or our narrowness puts us into that position, we are giving up our control of the situation and allowing ourselves to act in automatic fashion in response to the demands of the method or means in question.

There may be little danger that those who have worked constructively in the development of mental tests will fail to realize the limitations of them or be content with the use of them for its own sake. Probably, also, most administrators will ask how the tests can help in solving certain pressing problems. There is need, however, for more than this. Now that the tests have been developed to the point where we can say positively that they do serve with considerable success their immediate purpose of distinguishing groups of children on the score of differences in intelligence, it is time to review constructively our whole theory of educational organization with respect to the individual child.

Mental tests distinguish individuals in a new way. They give us information we have never had before, in a reliable form, about the status of any given child. They put us, therefore, in a new position with respect to our treatment of individual children. Accordingly, it is well to make sure that we know what we want to do for the children whom we can thus more effectively single out for special treatment.

The movement to adjust the school to the needs of individual children has a history of some length and much interest. In Ameri-

can schools individual instruction gave way to class instruction as a matter of practical necessity. We could not teach all the children economically until we had developed the technique of class teaching. Not long after the modern scheme of grading was established, it became clear that it had led to various evils and injustices. Since then, many schemes have been proposed and tried for handling large numbers of children without sacrificing the individual to the mass. Some of these are administrative schemes—plans for the grouping of children for purposes of grading and promotion, such as the so-called "Cambridge plan." Some involve the formation of special classes and the hiring of special teachers for work with selected individuals, as, for example, in the Batavia system. Some are schemes of method, such as the Courtis Practice Tests in Arithmetic. From one point of view, mental tests, as well as subject-matter tests, may be considered as new means for accomplishing the end for which all these other plans have been devised, namely, the individualization of instruction. If such plans as have been heretofore proposed were but external and limited in their application, we are now in a position to give them new and more fruitful trial. And just because we have a new means for individualizing instruction, we ought to ask again what we want to accomplish by it and what is the best way to do it.

We want to know more about children as individuals in order that we may deal with them as individuals. But that is not an end in itself, for individual treatment is just one mode of achieving the purposes of education and may be variously combined with treatment by groups. Individual treatment must itself be seen as a means to an end.

Furthermore, instruction is but one phase of education, and there is always the possibility that a new means for the improvement of instruction may lead to an overemphasis on intellectual development as compared with physical development or with moral and social development. There are real dangers here which ought now and then to be re-stated with fresh emphasis.

The ideal of complete development for every individual up to the limit of his capacities is extremely attractive. In general, also, it is probably a safe guide for practical effort if it be supplemented

by the notion that individual development must be in accordance with a definite plan which excludes some possibilities by the very fact of choice of others. William James made clear, in a famous passage, the necessity for choosing the self one wants to be. Within the limits of such a choice (which can not, of course, be made at once or very early in life), we ought to try to give every individual the chance to develop to his full stature. There are plenty of external limitations to this effort, for poverty, disease, and injustice will set at naught much that education attempts to do for children. All the more, therefore, should the school attempt to give each child his full chance. But we must remember to take the individual in his wholeness. Just now, I believe, there is real need for emphasis on physical development, for although some schools have learned how to watch bodily growth and adjust instruction to it, there is a general tendency to drift into fads of physical education rather than to safeguard health by simple means and allow time and space for natural growth. There is need, also, for renewed insistence on the importance of social and moral development—that maturing of character in the give-and-take of group enterprises, on the playground and elsewhere, for which no amount of book-work can be substituted.

All this, I realize, only states in dogmatic fashion what has been said more amply and convincingly by many others. G. Stanley Hall long ago warned us against precocity and a lopsided intellectual development. John Dewey has led a generation of teachers in their effort to manage school work so as to favor moral and social growth in children. The whole vocational-guidance movement is based on the assumption that each of us must make progressive discovery of the kind of person he wants to become. It would be useless to re-state these positions if it were not for the danger that mental tests will lead to new and uncritical attempts to achieve individual development on a partial view of what individual development means.

I have observed especially a tendency to assume that the only right and possible thing to do for bright children is to promote them rapidly through the grades. Heretofore, this has usually been done by "skipping," or at times by grouping children in rapid-

advancement classes. It has been done in the main on the basis of the ability of the children in question to cover quickly the work laid down in the course of study. Every practical schoolman knows that it has often led to disaster—that the child who has skipped a grade or done the work of two grades in one year has failed later in his course or broken down as the result of pushing. Mental tests are likely to help in avoiding that sort of failure, for they will enable principals to distinguish the children who are merely bright in the mechanics of school work from those who are fundamentally superior in intellectual ability. It does not necessarily follow, however, that there is nothing to do with a bright child, even if we are assured that he is genuinely of superior mentality, except to promote him rapidly through the grades. Here is a practical issue of administration in the elementary schools which the development of mental tests ought to bring squarely before us: Is rapid advancement for the mentally superior so generally desirable as to justify the formation of rapid-advancement classes or other schemes for putting these children through school faster than their fellows; or are there other and better ways of dealing with them?

An administrative scheme usually leads to an effort to make the machinery move. If classes for rapid advancement are formed, principals, teachers, and parents will unite to see that children are put into them. This will lead, I believe, even with the use of mental tests, to unfortunate results. In the first place, mental superiority will be used as a ground for grouping children without sufficient reference to physical development and social maturity. In the second place, many children will be pushed forward through the course of study when what they ought to have is an enrichment and differentiation of school work.

Suppose a class of mentally superior children has been selected on the basis of school marks and mental tests. Among them there may be many children who are big and strong and socially mature. There is at least some evidence to show that mental superiority goes pretty generally with physical superiority. Others, however, will not be well grown nor well developed in their powers of leadership or of cooperation. What does such a group need? Does it need the chance to go through the common branches of two grades

in a single year, or does it need, rather, shorter periods and more effective methods of drill and thus a saving of time for wider reading, dramatization, manual work, outdoor play, and other interesting and really educative enterprises, carried on by groups and largely in the form of projects?

There is no doubt that some children can stand being advanced rapidly through the grades, that they need to catch up with children of their own stage of development, or ought to be grouped with children chronologically older than themselves. To deal, however, with all children of proved mental superiority as if rapid promotion were the only way to deal with them is to confess poverty of resources and ingenuity. The whole child ought to be taken into account. More than that, natural social groupings ought to be taken into account. To select certain children for rapid advancement and to push them ahead of their fellows is not necessarily good for them, for the group they leave, or for the group they join. There is no evidence that pupils who enter high school or college young do not do well in their studies or that they get into disciplinary difficulties. Indeed, I have myself shown (Youth and the Dean, *Harvard Graduates' Magazine*, June, 1913) that the younger a man is when admitted to Harvard College, the greater is the likelihood that he will do well in his studies and keep out of trouble. It can not be said, however, that every boy or girl who is capable of saving time in his education by rapid promotion ought to be allowed to do so. Something should be said for normality. Health, companionship, and happy participation in the activities of his companions are considerations which should all be taken into account in dealing with every individual case. Education is a means whereby the individual may have full development among his fellows and for the common good. No short-sighted view of what individual development means should lead us to separate a bright child from the companions with whom he can be happiest and from whom he can learn most through common work and play.

It is true that those who go into the professions are often forced, in this country, to spend too many years securing an education. That is a problem in the adjustment of our scheme of education to the civilization it is serving. We ought not to conclude

that our program is properly outlined and that the thing to do is to hurry the bright ones through it while those of average power or less go on more slowly. Nature has a program in the development of children of which we must also take account, and it may be far better to curtail or telescope the higher stages of education, which come after natural development is more nearly completed, than to run the dangers of a forced pace during the earlier years.

Undoubtedly, children of superior mentality ought not to waste their time in the classroom while the teacher is struggling with the difficulties of duller minds. They ought to go through the minimum essentials at the faster rate of which they are capable. But before we assume that they ought on that account to be encouraged to complete their work in the grades and in the high school in less than the usual time, we ought at least to experiment with the plan of allowing them, instead, to use the time they save on school routine in freer, happier, and more rewarding ways.

This article is not an "attack" on classes for gifted children. There is ample evidence that gifted children can now be selected with satisfactory accuracy. It has been proved that they can be grouped for special treatment to their general advantage. What has not been proved as yet is the value, in any large administrative policy for handling classes of the gifted, of the element of rapid advancement. This article is but a "word of warning" on that score, from one who is not an expert in testing and who has had no part in the recent highly valuable experimentation in the treatment of children of superior mentality.

CHAPTER II

THE GROUP INTELLIGENCE TESTING PROGRAM OF THE DETROIT PUBLIC SCHOOLS

WARREN K. LAYTON
Psychological Clinic, Detroit Public Schools

There has been maintained in Detroit, for about ten years, a system of special classes for backward children and from time to time other units have been added, so that at present there is a department of special education equipped to care for pupils who, for any reason, do not progress properly in the regular grades. The Psychological Clinic, one of the earliest of these units to develop, is the agency through which transfers to the various special classes are effected. This clinic has had a rapid, but very solid, growth and enjoys the confidence and the support of the teachers and principals to a degree unusual in American cities. There are on the staff of the Clinic eleven trained psychological examiners and four social workers, all of whom give their full time to the work of the Clinic, and the Clinic has also its own physician.

Prior to the war, the service of the Psychological Clinic was rendered, of course, entirely through individual tests. The successful development of group tests of general intelligence in the United States Army in 1917 and 1918 and the adoption of the group method by hundreds of school systems is now an old story. Owing to its well organized psychological facilities, and especially owing to the progressive attitude of the Detroit teaching public, the inauguration of group mental tests in this city was brought about promptly. It is not the purpose of the writer to give a detailed account of all of the work that has been done in this field in Detroit, but rather to mention a few of the most important phases of the work and to present a statement of the organization and administration of the testing.

The studies of elimination and retardation of the past few years and the discovery by psychologists of wide differences in native ability among pupils have led many school people to the conclusion that education could be made much more effective if there were available a means of classifying pupils on the basis of mental ability, and with this in view many experiments have been and are still being carried on in various cities. In Detroit it was believed that to give the new plan of classification a fair trial it would be wise to classify by means of a group test all pupils entering school for the first time, and then to maintain intact the divisions thus formed so far as possible throughout the six years of the elementary course. The plan is to adjust the education of these groups of children of different mental levels entirely through the curriculum and the methods of teaching rather than to provide a scheme whereby the most capable pupils complete the course in less time. Briefly, our plan is this: for the "average" ("Y") group, comprising the middle 60 percent of the pupils, the present course of study; for the "backward" ("Z") group, comprising the lowest 20 percent, a simplified course of study containing minimal essentials sufficient to pass the pupil from grade to grade, and for the "superior" ("X") group, comprising the 20 percent at the top, an enriched course of study. Thus, all pupils, except the few very backward ones who cannot keep up even with the "Z" group, should complete the six years of elementary education without repeating grades. The few "Z's" who fall by the wayside will, of course, be the candidates for the special classes for backward children. The many interesting educational problems raised by this new classification must be omitted from this discussion, save to mention enough to give the background for what follows.[1]

At the time our plans were being made, there were few group tests available which were suited to children six years of age. After careful study of the problem and some testing with available group scales, it was decided to construct a new test for our purpose. Dur-

[1]A study of the first year's results of our new classification is now in progress and an account will be given in a forthcoming number of the *Detroit Educational Bulletin*, prepared by Dr. Charles S. Berry, Director of Special Education, Detroit Public Schools.

ing the spring and summer of 1920, the Detroit First-Grade Intelligence Test was developed and perfected.[2]

The test consists of ten separate tests, as follows:

1. Information
2. Similarities
3. Memory
4. Absurdities
5. Comparisons
6. Relationships
7. Symmetries
8. Designs
9. Counting
10. Directions

Most of the material is presented through pictures. The test was given for the first time in September, 1920, to about 11,000 children then entering our B-1st (lower first) grade and is now given regularly to all children entering the first grade. About 80 percent of these children attend the kindergarten, so it is possible for us to test them just before they leave the kindergarten and thus have the ratings in the hands of the schools at an early date. The examining is done by a corps of kindergarten teachers who have been trained for the work in special courses offered in Detroit Teachers' College by a member of the Clinic staff. The time required for the examining is about a week, and it takes ten days additional to score the papers and prepare typewritten lists of the results. A perfect score in the revised Detroit First-Grade Test is fifty points and letter ratings are assigned in accordance with the outline presented in the following table:

DETROIT FIRST-GRADE INTELLIGENCE TEST: RANGE OF POINTS FOR LETTER RATINGS

Score	Percent	Rating
0 – 12	8	E
13 – 17	12	D
18 – 23	18	C–
24 – 30	24	C
31 – 35	18	C+
36 – 39	12	B
40 – 50	8	A

The "A" and "B" pupils who constitute the highest 20 percent are recommended for the "X" group, the "C+", "C", and

[2]The test as originally constructed contained fifteen separate tests, five of which were dropped in the course of our first revision. The test as used at present, known as the Detroit First-Grade Intelligence Test, First Revision, is distributed by the World Book Co., Yonkers-on-the-Hudson, New York, and Chicago. Copyright, 1920, by Anna M. Engel.

"C—" pupils for the "Y" group, and the "D" and "E" pupils for the "Z" group. The score is not adjusted on an age basis, as most of the pupils entering Grade B-1 are homogeneous as to age. The highest score thus far recorded is 48 and the lowest 0. The first and third quartiles are 19 and 34, respectively, and the mid-score is 27 (true median, 27.59). The results thus far obtained indicate that this test classifies pupils from 6 to 7½ years of age with reasonable accuracy. Beyond this age it is not recommended. It is easy to administer, as the directions have been reduced to a minimum, and it requires no paraphernalia whatever. The time required for the test is from twenty to thirty-five minutes, according to nationality and home environment of the pupils tested. It is generally unwise to include more than ten or twelve children in a group.

Since September 1, 1920, the testing of B-1st pupils has constituted about 40 percent of our work with the group tests. Thus the testing of beginners in school is one of the most important functions of the group examining, as it should be.

Group tests, secondly, are given to pupils who are two years or more over-age for their grade, and to those who are persistently backward in their school work, to be followed later by individual tests of those making the lowest scores, and the subsequent transfer of some of these pupils to special classes. This examining is done in all elementary schools. Priority of this examining is decided, in part, by the availability of space for special classes in different parts of the city.

Group tests, thirdly, are given to children who are candidates for entrance to Special Advanced Classes, where there is an enriched curriculum suited to the requirements of unusually gifted children. These classes are now maintained in the 7th and 8th grades and are located at several convenient centers. Provisional candidates for the Special Advanced Department are chosen, of course, from the upper 6th grade and must be recommended by their teacher and principal. They must be either at grade or accelerated for their chronological age and must be marked either 1 or 2 for their school work (Detroit pupils are marked on a scale of 1 to 4). We then administer two group tests to these children and select for trans-

fer to the Special Advanced Department only those pupils whose scores are within the highest 10 percent in both tests. Since this method of selection has been used, the teachers in this department all report that the children are definitely of superior mentality and that they practically always make good in their classes.

The examining thus far outlined is done at the initiative of the Department of Special Education of which the Clinic, as has been said, is a component. Regular requests for group tests originating in the central offices of administration are for the examination of all new teachers and substitute teachers and of applicants for clerical positions in the offices of the Board of Education. Of more interest, perhaps, is the examining which is done at the request of the schools themselves, for the purpose of classifying pupils on the basis of mental ability. Thus far more than 10,000 children have been given group tests with this classification in view, always at the direct request of the principals of the schools. Four of the five intermediate schools have had their entire memberships examined. Requests for group tests in the senior high schools concern usually pupils in the A-12th grade, who are soon to be graduated, and who will be likely to require an intelligence rating in their entrance credentials when they enter the university. Four of the nine senior high schools have requested group tests of 9th and 10th grade pupils, for the purpose of assignment to sections in English and other subjects, and, in two instances, for assignments to home rooms. Two senior high schools have had their entire memberships examined. Eight elementary schools have had their entire memberships examined.

We have had a number of requests from the Department of Research for group tests where the scores are desired as a basis for important experimental investigations. Two such cases have been the examination of about 550 children in one elementary school and 300 in two others, to provide groups of like mentality for two experiments, one in reading and the other in measuring the effects of moving picture instruction. Recently we have examined about 400 high-school pupils as a basis for an extensive experiment in supervised study.

It is difficult to know just what is the best method of interpreting group test scores for the use of principals and teachers. At present we are using letter ratings for each test, similar to the plan used in the U. S. Army and corresponding to our own scheme adopted for the first-grade classification. Our plan is to tabulate the numerical scores of a given age group and then to assign the letter ratings in such a way that the highest 8 percent of the pupils are rated "A", the next 12 percent "B", etc., according to the outline presented in the table above. We never make these letter ratings until we have as many as three hundred unselected cases for a given age. The advantage of this plan is that it furnishes a basis of comparability for pupils of different ages. Of course, the different tests which we use vary somewhat in details, but not in their general nature. A six-year-old pupil who is rated "A" resembles a twelve-year-old pupil who is rated "A" in that each is among the highest 8 percent of his age group in intelligence.

The tests which we use regularly are as follows:[3] in Grade B-1, the Detroit First-Grade Intelligence Test; in Grades A-1 to A-4, a special test adapted for Detroit from the Army Beta, known as Test "X"; in Grades 5 and 6, a special Detroit test (Detroit Army Test) adapted from the well-known Army Alpha; in the intermediate school, the Terman Group Test, and in the senior high school and for the examination of teachers and other adults, the Army Alpha. All tests are given by the Clinic staff, and scored in the offices of the Clinic. This is done for several reasons, the most important being that the necessary uniformity of the examining and scoring procedure is insured when the work is in the hands of one trained staff. Another reason is that the group intelligence tests, themselves novel in character and differing ma-

[3]The tests named above are those which we are using regularly during the present year. We have made some use of other tests, as follows: the Pressey Primer Scale for the examination of pupils in the primary grades; Whipple's Group Tests for Grammar Grades in examining special advanced candidates and the National Intelligence Test, Scales "A" and "B," in grades three to eight. Doubtless some of these and others will be used again. We feel that the important thing is the use which is made of the test results rather than the specific test administered, though the latter is important. We tried to select primarily a test which gives the proper score distributions, but we are obliged to give some consideration, also, to such factors as length of time required for giving the test, time involved in the scoring and reporting, procedure, and also expense.

terially from the usual schoolroom tasks, appear to attract somewhat better performance from the pupils when administered by a stranger. This does not mean to say that the tests might not be given as well by the teachers—which might easily be the case—but simply that the uniform procedure and the elimination as far as possible of the personal element, both so desirable in work of this sort, can best be secured by using specially trained examiners. With the group testing in the hands of the teachers, themselves, there would be lacking the facilities for making the proper statistical interpretations based on a large number of cases, and for making letter ratings and the like, all of which is quite important.

In this connection the question has been raised: might not our system of group intelligence testing, apparently confined to one agency of the schools, operate to keep the benefits of the tests away from some interested teachers and principals? This is a misapprehension which cannot be removed too soon. So far as our facilities permit, with the exception already noted, we do any examining requested by any school where the principal and teachers wish to make use of the results. By this arrangement it is believed that in the long run the testing will be much more valuable. Rendering psychological examining service in response to requests in a school system containing 150,000 pupils is a task of some magnitude and it challenges the best efforts of our staffs. However, it is our earnest desire that our work shall not be limited to the extent that we become merely an examining agency. Thus we are receiving an increasing number of requests from the schools for specific recommendations as to placement of pupils. We wish to develop this phase of our work to a point where we can, by our recommendations, bring about in the different classes as nearly as possible uniform mental levels. This will not, of course, bring us into conflict with the function of the individual psychological test, which is an instrument for diagnosis while the group test is an instrument for classification. But we wish this development to occur in response to a need rather than as a consequence of an executive order. It would be possible for the Superintendent of Schools to direct that all pupils in the elementary and high schools should be given a mental test once annually. Many obvious advantages would accrue from such an arrangement and it is probably quite

true that there is a tendency toward just such a situation, as has recently been noted by Professor Terman. We feel that our plan of giving the tests (with the exception of grade B-1) upon the request of the schools is more satisfactory than a compulsory arrangement. To indicate the interest shown by the school people, it may be mentioned that in the ten months between September, 1920, and June, 1921, 58,000 individuals were given group tests in Detroit. As this is written (November 18, 1921,) we have exceeded 20,000 this year.

The group tests of intelligence have been developed in response to a need for some means of ascertaining the fundamental individual differences in native ability which we now know to be among the most striking phenomena of mental life, and of using this information for a better basis of classification of individuals for instruction or for other purposes. The administration of the tests constitutes an effort to be useful to the teachers and others in charge of the training of the pupils whose gifted or limited mentalities form the raw material of the educative process. We believe that in their proper field group intelligence tests can be a very great help to any teacher in any school; they will solve many maladjustments at once and save much of the labor and discouragement always brought on when pupils are attempting to do work that is unsuited to their ability. The group test is not, however, an instrument for the *analysis* of the difficulties of individual pupils: it is an instrument of *classification;* it establishes the intelligence-group to which the pupil will almost surely be found to belong and in which there is every reason to believe, other things being equal, that he will do his best work. For the backward pupil who makes the "E", or lowest, rating by the group test, or the pupil of unstable or erratic temperament, the group test is not enough. Here a study of the case is of the utmost importance, and this study should take the form of an individual test, accompanied by a medical examination and a social history.

We are gratified by the constant and substantial increase in the number of group mental tests in Detroit because it reflects a great interest on the part of the teachers and principals and because the teaching public shows an earnest desire to make use of the test results. Such a genuine interest, it is a pleasure to serve.

CHAPTER III

THE USE OF INTELLIGENCE TESTS IN THE CLASSIFICATION OF PUPILS IN THE PUBLIC SCHOOLS OF JACKSON, MICHIGAN

HELEN DAVIS
Director of Measurements and Special Education, Jackson, Michigan

There are numerous school systems, apparently, in which more or less systematic use has been made of intelligence tests, but in which the scores obtained from these tests have not been put to the fullest possible use for the improvement of organization, placement, and instruction. Naturally, the extent to which reclassification can be effected on the basis of test results is dependent upon the general lay-out of the system in question, the distribution of ability in its population, its financial resources, the availability of classrooms and teachers, and many other factors. It is probable, indeed, that no scheme could be laid down in detail that would fit any large number of school systems. Nevertheless, it has seemed likely that an account of the manner in which a plan of intelligence testing has been related to a system of special classes in one American city might prove helpful to those who are undertaking similar work in other cities of similar size and character.

THE GENERAL PLAN OF SCHOOL ORGANIZATION AT JACKSON

1. The Regular Classes

Jackson is an industrial city of approximately 50,000 population and enrolls in its public schools some 7,000 children. The elementary schools include the kindergartens and grades low one through high six. Two intermediate schools, one on either side of the city, include grades seven, eight, and nine, while the single central high school includes grades ten, eleven, and twelve. The regular grades of the system, therefore, conform to the familiar six-three-three type of organization.

2. The Special Classes

There are at present seven types of special classes in Jackson (eight if we count the upper and the lower auxiliary classes as distinct types). So far as my information goes, I judge that Jackson, under the progressive leadership of Superintendent Marsh, has gone farther than most cities of its size in the elaboration of its system of special classes; at least there are numerous systems larger than ours in which special provision for atypical pupils is limited to a few ungraded classes and perhaps provision for individual promotions of gifted children.

The special classes for the blind (conservation of vision classes), for the deaf, and for the anemic are in the main recruited through other departments than the Department of Measurements[1] and through other agencies than intelligence tests. For this reason no further reference will be made to them or to their work in this discussion.

The remaining special classes comprise four types, each of which demands explanation. The facts concerning these classes are for convenience, summarized in Table 1; they are set forth in more detail in what follows.

a. The "Ungraded Classes." There are ungraded rooms on each side of the city to which children are sent who are known to be definitely feeble-minded. These rooms draw their pupils from any of the elementary grades and even from the intermediate schools, though in practice children of this mental caliber are rarely found above the fourth or fifth grade of the regular classes. As a rule, the pupils assigned to a room of this type complete their school careers within its walls and are not returned to the regular classes.

[1] The Department of Measurements was organized at Jackson in the fall of 1920. It ought to be added that several types of special classes were in operation in the system before the establishment of the Department. The work of the Department, however, has placed the selection of pupils for these classes on a more systematic and scientific basis and has also led to the establishment of other types of special classes, notably the auxiliary classes. Readers who are interested in the operation of such departments and in their relation to other branches of the school system will find an account of our experiences in Jackson under the title "Some Problems Arising in the Administration of a Department of Measurements," *Jour. of Educ. Research*, Jan., 1922, pp. 1-20.

TABLE I.—ORGANIZATION AND RECRUITMENT OF SPECIAL CLASSES IN THE PUBLIC SCHOOLS OF JACKSON, MICHIGAN

Name	Grades	C. A.	M. A.	I. Q.	Tests Used	Criteria for Admission	Curriculum	Length of Stay	Number Enrolled	Number Teachers
A. Ungraded	Kg. and above	Any age	4½-10	Below 70	Binet for final. National and Haggerty, Delta 1 for preliminary	Definitely feeble-minded	Academic work in so far as the child is capable. Development of motor control. Pre-industrial training. Training for citizenship.	Until compulsory age-limit is reached	108	6
B. Opportunity	5 and 6	14 and above	10 and above	75 and above	National, A 1 Binet	Mental ability and school training equal to or better than that of a 5th-grade child	Academic work of 5th- and 6th-grade level, together with 7th-grade work in shop, gymnasium, cooking, etc.	Until 6th-grade work is completed.	80	2
C. Auxiliary 1. Upper	4, 5, 6	12 and above	9 and above	70-85	National, A 1 Binet	I. Q. between 70 and 85. Poor school record	Minimum essentials of the regular course given at a slower rate and by somewhat different methods.	Until ready for opportunity or regular classes	48	2
2. Lower	Kg.-3	below 12	4½-9	70-85	National, A 1 Haggerty, Delta 1 Binet	I. Q. between 70 and 85. Poor school record. Occasional cases of normal intelligence and pedagogical disability.		Until ready for upper auxiliary or regular grade	72	3
D. Speed	3B-6A		1½ yrs. or more accelerated	115 or above and M. A. equal to the grade	National Intelligence A 1 Whipple Group tests, Haggerty, Delta 1, Binet	85th P. C. or better in age group on group tests	Two semesters of the regular grade work completed in one	One semester. In rare cases two semesters	90	3

These pupils range in chronological age from 7½ to 16 years, in mental age from 4½ to 10 years. The course of study and the methods of instruction are similar to those prevailing in ungraded rooms generally in American school systems. The organization of this type of class is such, however, that the work is departmental in plan.

b. The "Opportunity Classes." The "Opportunity" rooms are located in each of the two intermediate schools. They are designed to meet the needs of pupils who are "over-age" (14 years and over), of a fair degree of mental ability (mental age, 10 years and above), but who have become so retarded pedagogically as to be doing only fifth- or sixth-grade work. The plan is to give these pupils instruction suited to their needs and at the same time to give them an opportunity to associate with children more nearly their own age. Their course of study includes materials and subjects characteristic of the grades mentioned, but in addition they may earn credit in some regular seventh-grade subjects, such as shop, gymnasium, cooking, printing, sewing etc. It is hoped that by this course of study their interest in school work will be prolonged a few years more and that they will be better equipped to meet the demands of life after they have left school.

c.1. The "Upper Auxiliary Classes." The operation of ungraded classes in any school system soon reveals the needs of a group of pupils who are not sufficiently inferior mentally to be placed in these ungraded classes, but who are at the same time not sufficiently capable mentally to keep the pace of the regular classes. In our system the needs of this group of so-called "slow-dull" pupils are being met by the establishment of another variety of special class. These classes, to which the term "auxiliary" has been applied, were put in operation in September, 1921. They may be regarded as an extension downward of the Opportunity Classes just described. The "Upper Auxiliary Classes" are composed of pupils 12 years old and above chronologically and about 9 years or more old mentally. Their intelligence quotients are, then, between 70 and 85. As a group they are characterized, as might be expected, by poor school records; in fact, 80 percent of them have

failed from one to four times and 50 percent of them have been conditioned from one to three times. After transfer to the auxiliary rooms they carry on work of the fourth, fifth, and sixth grades, but stripped to the "minimal essentials" and conducted at a slower pace and by somewhat different methods than in the regular grades of the same scholastic level. At present there are in operation two rooms of this sort, enrolling 48 pupils.

c.2. The "Lower Auxiliary Classes." These classes are composed of pupils below 12 years of age chronologically and under 9 years of age mentally. Their intelligence quotients, like those of the pupils in the upper auxiliaries, range from 70 to 85.[2] Here again the school records are poor; 60 percent have failed from one to four times and 16 of the 72 pupils now in the three rooms of this type have been conditioned from one to five times. The work undertaken ranges from that of the kindergarten to the third grade, and, as in the upper auxiliaries, is limited to the essentials and conducted at a slower pace and by somewhat different methods from those prevailing in the regular grades.

It may be noted in this connection that the classification of pupils by intelligence tests has given new emphasis to the demand for a revision of the course of study and methods of instruction to meet the needs of pupils whose intelligence differs so clearly from that of the "average" pupil. In Jackson we are trying to devise new ways of teaching the essentials to these duller pupils. Clay work, games, tools, charts of individual accomplishment, projects and other devices are being used to stimulate interest, and monthly records of school work are being kept to indicate the progress attained under these modified conditions. Similar work is under

---

[2] On account of certain geographical difficulties in the transfer of pupils to the ungraded classes, a few definitely feeble-minded pupils have been temporarily placed in the lower auxiliaries, but these pupils are to be transferred again to ungraded rooms as soon as the difficulties of transportation can be met.

There are also two or three special cases of children who are normal in mental ability but handicapped by a particular pedagogical disability, notably the inability to read, who have been put into the lower auxiliary classes where it is hoped that the modified procedure and the opportunities for individual instruction will enable them to bring up their performance to the level where they can resume regular grade work.

way in many other cities, and it is not too much to hope that in time there will emerge a satisfactory program with modified textbooks, modified methods and modified subject matter that will effect far-reaching improvement in our training of these pupils.

d. The "Speed Classes." The so-called "Speed Classes" in Jackson are at present three in number, with an enrollment of 90 pupils. The rooms are situated on either side of the city and are designed, as their name implies, for pupils of superior ability and attainment. Pupils are admitted to these rooms from the upper second through the upper fifth grades. Generally they remain in the speed room for one semester where they do the work of two regular semesters and are then returned to the regular classes; occasionally, exceptionally capable pupils are allowed to remain two semesters in the speed room, *i.e.*, to accomplish two years' work in one year. The selection of pupils for these rooms is mainly effected by the use of group intelligence tests.[3]

The criterion for admission is the attainment of at least the 85th percentile in their *age* group (due regard being taken for the proper relation between chronological age and grade); this means that the pupils selected must have equalled or exceeded the median score for children two years their senior, or in other words, that they must be two years or more accelerated mentally.[4] The opinions of the teachers of the pupils provisionally chosen by the group tests are always solicited. Usually these opinions confirm the results of the intelligence tests. If, however, the child's classroom performance does not seem to warrant his transfer to the speed room, an individual examination by the Stanford Revision of the Binet test is usually made. In cases where the child's group test record is unusually good (90th percentile or better), but the teacher's judgment is adverse to the transfer, the elementary supervisor is usually consulted with regard to the professional skill and critical

[3] The group intelligence tests employed for the selection of candidates for the speed rooms have been the National Intelligence Tests, the Whipple Group Tests for Grammar Grades, and the Haggerty Delta 1 (for the younger pupils). Recent experience shows that the use of two such group tests insures a much more reliable selection.

[4] This is the criterion with the group test; with the Binet some pupils have been selected who were accelerated only one and a half years.

judgment of the teacher in question; if then it turns out that the teacher's standards are unusually high or her tendency is to place undue emphasis upon drill and the mechanics of subject matter, the child has been given a trial in the speed room without further examination of his intelligence.

The Intelligence Testing Program at Jackson

1. Group Intelligence Testing

Before the Department of Measurements was created, pupils had been selected for the special classes on the basis of the teachers' estimates only, with the exception of the ungraded room in which case pupils adjudged to be feeble-minded had been referred for a Binet test to the teacher of this room, who then admitted the most needy. The policy of the Department of Measurements was to utilize from the start the system of special classes then in operation, but to put the selection of children for these classes upon a more comprehensive and systematic basis. To this end the National Intelligence Test, Scale A, Form 1, was given at the outset to all pupils from the low-third through the high-sixth grades, inclusive. By giving careful preliminary instructions to the teachers, over 2500 pupils were tested simultaneously. The test blanks were then scored by the teachers, and forwarded to the office of the Department, where they were re-scored, and where distributions were made, grade medians for the city and for each school were computed, and age percentiles were determined.

We quote here a paragraph of explanation concerning these percentiles that has appeared elsewhere.[5]

"Since most of the information concerning the location of children in the grades is familiar to teachers and supervisors in terms of mental age, it was felt worth while to translate the scores of the National tests into 'Jackson mental ages.' This was accomplished by regarding the median score of pupils of each age group as the standard score for the mental age as well as the chronological

[5] G. M. Whipple. "The National Intelligence Tests." *Jour. Educ. Research.* 4: June, 1921, pp. 28-29.

age of the group in question. Thus, all pupils aged eight (over eighth birthday and under ninth birthday) were distributed in such a way as to locate the median and all the other deciles, and this median was regarded as indicating a mental age of 8½ years. The medians for 9½, 10½, and 11½ years were located similarly and points midway between these medians were taken as the scores indicative of mental ages of exactly 9, 10, and 11 years. The amount of overlapping was shown graphically by the percentile chart, and this chart became directly useful in locating pupils of any desired degree of deviation from the standard adopted for a given grade or group. Thus, pupils were drawn off for consideration in connection with ungraded classes and speed classes, for double promotions, etc."

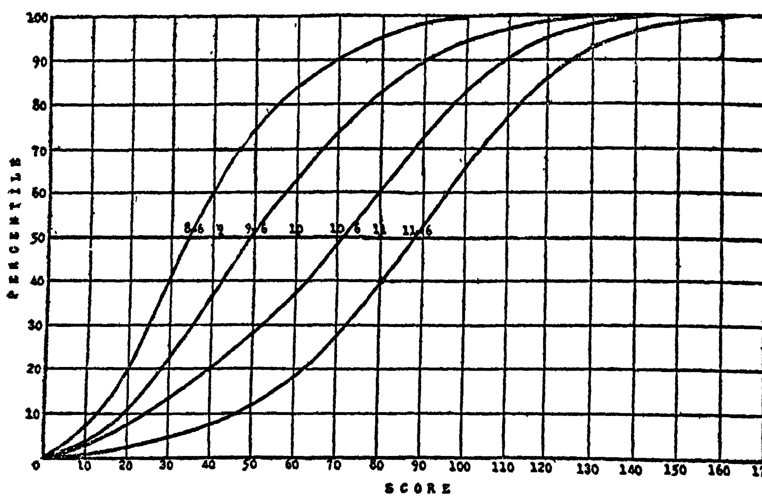


FIGURE 1.—PERCENTILE CHART FOR THE NATIONAL INTELLIGENCE TEST, SCALE A, FORM I (JACKSON, MICHIGAN)

"It will be understood that in this chart each of the four age-groups of pupils has been reduced to a theoretical 100 pupils. The figures on the base line are the scores obtained; the figures on the vertical lines are the numbers of pupils in order of excellence. Thus, in the group aged 9 years (median age approximately 9 years, 6 months) the twentieth pupil in a hundred counting from the poorest pupil scores 28, the fiftieth (or median) pupil scores 49, the eightieth pupil scores 87, etc. Or, again, 25 percent of the 8:6 group score as high as the median of the 9:6 group, etc."

On the basis of these scores and computations, then, the task of placing pupils in the special classes appropriate to their needs was begun. It is perhaps not necessary to explain that individual examinations were given to many pupils; in fact, invariably given before transfer to the ungraded rooms, though the group tests even here were of decided usefulness, since all pupils whose group test scores ranked at the tenth percentile or lower were at once considered prospective candidates for the ungraded room.

Recently the National Intelligence Test, Scale A, Form 2, has been given to all pupils in the high-sixth grade preparatory to classification in the entering grade, 7B, of the intermediate schools. The pupils attaining the higher scores will be permitted a certain freedom of election denied the other pupils.

In addition to the National Intelligence Tests, the Whipple Group Tests for Grammar Grades have proved useful for selecting gifted pupils from the fourth grade and the fifth grade as candidates for the speed classes (these tests were especially designed for the selection of gifted pupils).[6]

Since the need of early classification soon becomes apparent, once any systematic classification is attempted, we have been experimenting with group intelligence tests for primary and kindergarten children. An elaborate comparative study of the merits of the Dearborn, the Haggerty Delta 1, the Kingsbury, the Otis Primary, and the Detroit First-Grade Tests was conducted at Jackson in the spring of 1921 by Miss Margaret V. Cobb, then Secretary of the Bureau of Mental Tests and Measurements of the University of Michigan.[7]

2. Individual Intelligence Testing

From the outset the Department of Measurements has continued the work of Binet testing that had been started prior to the creation of the Department. As already explained, the Binet test is used to confirm the assignments of all ungraded pupils, and of

---

[6] A special report upon the validity of these tests for this purpose will appear in an early number of the *Journal of Educational Research.*

[7] The results of this study are to appear in a doctorate thesis by Miss Cobb.

all or nearly all the doubtful assignments of pupils destined for the opportunity, the auxiliary, and the speed classes. A considerable portion of the Director's time is thus engaged in this work of individual examining.

Admission to the First Grade. In addition to this work of checking the results of the group testing, there has been developed at Jackson a plan for using the Stanford Revision on a much more elaborate scale for controlling the admission of pupils from the kindergarten to the low-first grade. In November and December, 1920, all kindergarten teachers in the city were given a fairly rigorous course of instruction in the use of the Stanford Revision. Before the opening of the second semester (spring of 1921), these kindergarten teachers had given individual examinations to 362 children and the Director had tested 58 others, so that we knew the mental age and the intelligence quotient of 420 prospective candidates for admission to the 1B grade.

Under the system prevailing prior to this experiment, any child who would be six years old chronologically before the end of May (that is, 5:8 at the opening of the second semester) might be admitted to the 1B grade. There is fairly conclusive evidence that children whose mental age is under six years are not likely to do satisfactory work in the first grade, but it was deemed expedient, under the conditions prevailing at Jackson, to set the standard for that particular semester at 5:8 mental age. In addition, it was provided that all children who at the beginning of the semester were 6½ years old *chronologically* might enter the first grade, regardless of their mental age. Of the 420 kindergarten children examined, 100 were held in the kindergarten on the basis of their test scores. Of this 100, 68 were more than 5:8 years chronologically and would, accordingly, have been admitted to the first grade under the old system. On the other hand, there were admitted to the first grade 50 children who were less than 5:8 years chronologically, and who would have been held in the kindergarten under the old system, but who tested 5:8 or better in mental

TABLE II.—RELATION OF MENTAL AGE TO SUCCESS IN THE LOW-FIRST GRADE AT JACKSON, MICHIGAN

(Spring Term, 1921, 277 entrants, excluding repeaters, foreigners, and transients)

Mental Age	6 or over		5:8 to 6:0		Below 5:8	
Outcome	Cases	Percent	Cases	Percent	Cases	Percent
Promoted	156	81.2	46	59.0	0	0
Conditioned	12	6.2	4	5.1	0	0
Failed	24*	12.5	28	35.9	7	100.
Total	192	99.9	78	100.0	7	100.0

*Of these 24, 10 were absent one month or more in all.

age (the mental ages ranged from 5:8 to 7:2, the I.Q.'s from 104 to 133.[8]

The results of this experiment in admission to the first grade are summarized in Table II, where it is evident, as others have already shown, that there exists a positive correspondence between mental age and success in the primary work. Eighty-one percent of those who had attained a mental age of six or more at entrance were promoted at the end of the semester, whereas only fifty-nine percent of those who had attained a mental age of from 5:8 to 6:0 were promoted, while all seven of the pupils whose mental age was less than 5:8 at entrance failed in their primary work. In the future it will be our policy to limit entrance to the first grade, in so far as feasible, to pupils who have attained a mental age of 6. In the second semester, however, on account of the smaller number of applicants and the desirability of keeping a reasonable balance between the number entering in the fall and the number entering in the spring, the mental age standard will of necessity be somewhat lower than 6 years.

[8] The youngest child admitted in this group was just five years old chronologically and just 5:8 mentally. That he was ready for first-grade work seems evident from the testimony of his teacher who reported later that he was "doing first-class work" and "better than some of the older ones."

In general, it may be said, the reaction of the first-grade teachers toward this method of admission has been most favorable, though a few of them are still reluctant to accept children less than 5:8 chronologically. The teacher was perhaps not speaking entirely in jest when she said that in addition to the intelligence test, the Department should "give them a performance test to see whether they can put on their rubbers and button their coats!"

In conclusion it may be said that the use of intelligence tests in the classification of pupils in this school system has received the hearty support of the teachers, that the pupils transferred to the special classes are happier and more successful in their work, and that the parents, once the purpose of the special classes has been explained and the children have had time to adjust themselves to the new conditions, are appreciative of the special provision that has been made for their children.

CHAPTER IV

MEASUREMENT OF THE ABILITIES AND ACHIEVEMENTS OF CHILDREN IN THE LOWER PRIMARY GRADES

AGNES L. ROGERS
Goucher College, Baltimore, Maryland

Once started, measurement in the lower-primary grades has advanced with considerable rapidity. It was remarkably late, however, in beginning. For this there was a variety of reasons. Prominent among them is the lack of agreement among educators concerning the earliest years of school life. Not only is there difference of opinion as to when a child should enter school, there is also still greater uncertainty and confusion of ideas as to the ideal course he should have after entrance.

In the face of such a lack of unanimity as to the specific objectives of the first school years, those equipped to measure mental products have avoided the labor of devising measuring rods for what might prove to be mere passing fancies or outworn fads of teachers of those years, rather than the permanent educational desiderata for children from four to eight. This, we admit, is an explanation rather than a good reason for the late beginning made, since nothing would contribute more to the definition of the objectives of lower-primary education than measurement intelligently applied. The clarification of the aims of high-school mathematics, consequent on measurement would suggest this and justify us in anticipating similar results.

A second cause for the present situation is the fact that those equipped with the training necessary for the construction of measuring instruments for mental abilities have generally had little experience with *young* children and naturally devoted their attention to the higher grades, and a third obvious reason for the paucity of work done was the intrinsic difficulty in devising suitable

tests for the youngest pupils. A new technique for group measurement is necessary in their case and the relative unfamiliarity of those trained in mental measurement with five and six-year-olds engenders doubt of the success that would attend attempts to measure their abilities or achievements.

The practicability of the application of group intelligence tests to men of low mentality and to illiterates in the U. S. Army, naturally hastened the construction of tests for pupils of six and seven. Already there are twelve group tests of general ability available for those years, and of these, norms for children of five have been established for one test, norms for children of six for six tests, and norms for children of seven for seven tests. Of group tests of achievement eleven tests are on the market, and three of these are standardized for the first grade and eight for the second.

Many of these measuring instruments are admittedly still in experimental form. Nevertheless even to-day, we have some proof of the predictive power of at least seven of them. They show, too, interesting improvements in technique of administration. Though much remains to be done, much has already been accomplished.

Content, Form, and Administration of Tests

In content the tests are pictorial. This, in itself, is a decided limitation. Individual examinations, such as the Binet-Simon Intelligence Scale in any of its revised forms, are undoubtedly more representative of a wide variety of abilities, notably linguistic and motor capacities. It has to be admitted, moreover, that linguistic abilities are paramount in importance for success with the customary elementary school curriculum. The ability to read is unquestionably the fundamental requirement for elementary school work, since m[illegible]y of many other subjects depends upon it. Those tests which [illegible]is important capacity are therefore of exceptional signifi[illegible]ictorial tests, as devised for little children, require comp[illegible]n of oral language, but they demand no ability to manipulate language. Indeed, it may be said with considerable justification that pictorial tests for children in the lower-primary grades weight far too heavily the mere comprehension and following of oral directions. There are differences of opinion as to the nature of

general intelligence, but whatever its constituent elements may be, it is certain that it is not such that it can be adequately gauged by just one type of mental performance. Success with each and every item in intelligence tests depends upon the ability of the individual child to take a group direction. This latter ability is largely affected by practice and in her work one teacher may seek to develop it much more than another. It follows that some process of equalizing opportunity in this respect is essential. Two methods are possible; the provision of fore-exercises, which might take the customary form used in testing older persons, or the application of a similar examination on a previous day. There is much to be said in favor of the latter method. Some, who have had experience in applying tests to children from six to eight, are of the opinion that in their case the adjustment to the test situation as such, can effect a greater improvement in scores than with children in higher grades. There is likewise good reason for preliminary training in the specific acts involved in the response made, but extrinsic to the particular abilities which are being probed. Such training could include the habituation of such responses as "Pencils up," "Pencils down," "Turn the page," in which there are great individual differences in the rate of work which might conceivably influence the scores and make impossible useful comparison with standards.

The reduction of the number of such specific responses is obviously desirable and the devising of scales which require but a single response, and that having only one possible interpretation, as in the Pressey Tests, is an important contribution. It represents a tremendous saving of the teacher's time in learning to give and score tests, and there can be no doubt whatever that it makes it much easier for the child of the mental age of five or six to sustain his attention. Where the tasks involved in the various tests of any examination require different kinds of reactions, confusion is apt to arise.

Another requisite on the content side of the tests needs to be mentioned. It is essential that concepts incidental to the abilities being measured, yet necessary for successful responses, should be verified as already established. For example, the making of digits

or letters of the alphabet or the comprehension of the meaning of zero are required by certain tests for children at the end of the first grade. It is necessary to make sure that mastery of these has been gained, otherwise we are not measuring the abilities intended to be measured, but something else.

Indeed, it has to be broadly affirmed that a fundamental desideratum of such pictorial tests for little children is that they must be adapted to their natural interests and experience-level. Certain pictorial tests can be extremely abstract in character and uninteresting to six-year-olds, and while it is an unattainable ideal perhaps, to expect any examination to demand no experience that any one child has not had, still existing tests show some noteworthy illustrations at variance with this ideal.

The very form of the tests demands the most meticulous care in the application of the facts and laws of mental development. The crucial problem after all is control of attention. If attention is not secured, intelligence cannot possibly be tapped. Sometimes the content or the method is such that tests fail to arouse the attention and interest of children. Invariably in testing we are careful to prevent the interference of such instincts as hunger and thirst. We give the tests at a time when these are unlikely to intrude and vitiate our results. It is equally essential that we should so control the stimulus presented to the child as to obviate other interfering tendencies. Thus, much experimentation is desirable on the ideal form of test. Should, for instance, the pamphlet-form be used at all, or is it almost impossible to control curiosity sufficiently to prevent children of six, in spite of directions to the contrary, from turning pages at inopportune moments? Again, what is the desirable spacing of pictures? Are not some of the existing tests too crowded, and consequently do we not have a dispersion of the child's attention rather than concentration on the task in hand? To take one illustration from one of the best existing tests, in the Kingsbury Group Intelligence Scale for the Primary Grades, is it not bad procedure, betraying ignorance of children of six, to have a two column arrangement in which, after completing the first column, the child is expected to begin at the top of the second and work down it? Is there not shown an almost uncon-

trollable tendency of six-year-olds to answer the items out of order and even to such a degree of distraction as to make them fail to grasp the group directions and merely respond according to their own undirected pre-dispositions to act towards such material? There is no room for question that too large an amount of material presented to the child has a bewildering and confusing effect, and the determination of the optimal number of different tasks we can present to the child of five or six for successive treatment is desirable. Existing tests vary greatly in merit as regards spacing of pictures, size of pictures, number presented, and clarity of printing. Unless these are controlled, we are in no better case than if we neglected to obviate noises, interruptions, or contrasting stimuli of any kind.

Another drawback attending the testing of young children which is usually absent at higher ages, is the untrained instinct of communication. This tendency is natural, and schools are more and more endeavoring to utilize it wisely, building upon it the mastery of the vernacular, the development of skill in drawing, and so forth. It is at this age almost impossible for some children to work independently. Contrary to the belief that the tendency to work together becomes stronger at adolescence, it would seem as if many children of this age habitually respond by seeing what others do, and find greater satisfaction in responding after seeing what another's response is. The obvious method of eliminating this is to seat children in such a way as to make communication impossible. None of the tests sufficiently emphasizes the care the examiner must exercise in seating children. Older children make known the fact that they cannot hear well or find the examiner's voice difficult to understand. The examiner of little children has to arrange the situation in advance of the test so as to find out for himself which children are experiencing difficulty in this way, and has to exercise judgment in discovering those children who are habitually dependent on others in their work.

Evaluation of Tests

A satisfactory beginning has been made in the evaluation of tests. Such a study as that of Holley is a more valuable contribu-

tion at the moment than the construction of a new test. At the present stage we require to find out much more concerning their comparative predictive power, and their relative convenience and reliability. Holley applied only one test, the Pressey Primer Scale, to children in the primary grades. The comparison with standards which his results afford in the case of that test is very useful. Further studies of this description are about to be published and will do much to advance knowledge in this field. We urgently need such systematic application and evaluation of existing group intelligence scales for the youngest children.

The problem of evaluation in their case is not so simple as at higher levels. Teachers' estimates and school marks are even more unreliable at these ages than later. Even if rating scales for these years are speedily devised, which may refine the judgments of teachers to an appreciable extent, this will still hold good. Much of the failure of mental tests at all levels can be traced to inadequate theory, and fortunately attention is now being concentrated on criteria for their validity. Increase in achievement from one age to the next and variations in achievement for children of the same age are now being supplemented as essential criteria by the power of such tests to discriminate adequately between two groups of children, one of notably superior capacity, the other of notably inferior mentality. The degree of correspondence found also between the results of group tests and individual examination of established trustworthiness, such as the Stanford Revision of the Binet-Simon Scale and the success of children in after years, are likewise valuable checks on the effectiveness of particular tests.

Achievement tests offer special problems from the standpoint of evaluation. In addition to the fact that we must have some guarantee that they do measure abilities that are worth fostering in school, it is essential that these tests should be in harmony with sound educational theory and practice. There is some likelihood of tests being published that do not meet these requirements. Opinion is greatly divided as to the content of the course for the first school year and perhaps it is an excellent thing that at this time so much experimentation is being carried on with an abundant variety of materials involving a correspondingly wide range of mental capacities.

Is there less room for difference of opinion on the second requirement? Such a test as Pressey's First Grade Vocabulary test has been criticized from this point of view. The effect of such an instrument might be to encourage the teaching of reading by developing word-getting rather than thought-getting. It may be answered that the occasional application of the test would work little harm and would be a useful index to the proficiency attained in comprehension. It is felt, notwithstanding, by many to be dangerous to place it in the hands of the teacher, because of its probable misuse. The scale announced by the Department of Research at Detroit certainly encourages a more valuable sort of reading ability.

Uses of Tests

Certain valuable studies have appeared in the course of the past year, which show the uses to which tests are now being put in the earliest school years. Notably Dickson has shown that if a child has a mental age of six he can do the work of the first grade, whereas if his mental age is less than that, he is found unable to cope with first-grade work. Evidence that the achievement of children in the primary grades is conditioned and limited by their mental maturity has likewise been presented by Arthur and by Haggerty.

Intelligence tests thus serve the important purpose of classifying children in accordance with capacity, which seems to be a necessary step even with children in the first school year. They prove equally useful as one factor in settling promotions. Buckingham has published facts that show that if a child has failed to attain the standard of attainment required for promotion, it is questionable whether the year's work should be repeated in all cases, and that promotion to a new teacher may give enough stimulus to make good the deficiency.

It is when coupled with tests of achievement that intelligence tests become most fruitful. Indeed, achievement of pupil or teacher can only be estimated fairly on the basis of such knowledge. The combination of intelligence and achievement tests also furthers diagnostic study and treatment of individual needs. Such investi-

gations as those of Anderson and Merton and of Zirbes indicate the large field which yet remains to be ploughed.

Tests of attainment may serve the additional purpose of measuring the efficiency of different methods of instruction and the relative merits of different courses of study. Theisen, for example, presents some evidence that children who have had kindergarten training show better results with reading in the first grade than do children who have not had that training. Such investigations enable us to evaluate more justly the kindergarten curriculum and methods and are fruitful of suggestions as to the kind of experience the child needs prior to learning to read.

In the future, investigations to determine a satisfactory course of study for the first school year will be made by their help; indeed, studies of this kind are now being made. For example, in primary education there is no greater need than an inventory of the specific habits and attitudes which we have a right to demand in normal children after a definite amount of time spent in school. The measurements of the important achievements represented by habit-forming will do much to concentrate attention on a most important aspect of education and one which is not only essential to success in social life, but also to success with later intellectual work. There is reason to believe that the fundamental habits of successful intellectual activity can be established much earlier than it has been customary to suppose. The fastening of the attention of the teacher on these rather than on subject matter, will bring excellent results and recognition of the gifts of those teachers who are exceptionally successful in this work is only their due. This may awaken, even in those neglectful of this branch of education, realization of the need for securing accomplishment in this respect, also. No such objectives have been specified in the past, and the teachers of five and six-year-olds would profit greatly if they were at hand.

This is but one phase of curriculum analysis of which no stage of education stands in greater need than the first few school years. At the moment the diversity of practice is great, and the only guide we have in the matter is common sense. There lies ahead of us the detailed study of achievements in order that standards may be laid down. Curriculum-making will not be the work of the psychologist

alone, but the psychologists's contribution of facts will give a basis for wise prescription in the matter. Only by determining accurately the actual accomplishments of children and their rate of progress can we arrive at curricula that can lay claim to being scientific.

Studies such as those of Packer on the vocabularies of first-grade readers and of Starch on the content of readers represent another side of quantitative investigation which will lead to scientific curricula for the primary grades.

Attention must also be turned to the making of rating scales for young children for those qualities of character for which no objective measuring rods exist and for which it is most unlikely that they will be forthcoming. These should be usable instruments that will refine and correct the teachers' judgments about pupils. They should cover those elements in character or personality which are essentially dynamic. Such scales are valuable in diagnosis of the causes of retardation and together with intelligence tests help greatly in locating sources of failure in school work.

The amount of retardation in the United States amounts roughly to over thirty percent and of this a substantial part can be traced back to the first grade. The discovery of the causes for this retardation should be the central business of departments of educational research. We may confidently expect that tests and scales for the earliest school years will loom larger in educational literature in coming years.

REFERENCES

Anderson and Merton. "Remedial work in reading." *Elementary School Journal.* May and June, 1920.

Arthur, Grace. "An application of intelligence tests to the problem of school retardation." *School and Society,* 10: 614-620.

Buckingham, B. R. "An experiment in promotion." *Journal of Educational Research.* 3: 1920. 326-335.

Dickson, V. E. "What first-grade children can do in school as related to what is shown by mental tests." *Journal of Educational Research.* 2: 1919. 475-480.

Holley, C. E. "Mental tests for school use." *University of Illinois Bulletin.* Vol. XVII. No. 28.

Packer, J. L. "The vocabularies of ten first readers." *Twentieth Yearbook of this Society.* 1921. Part I, Chapter 9.

Starch, Daniel. "The contents of readers." *Twentieth Yearbook of this Society.*" 1921. Part I, Chapter 10.

Theisen, W. W. "Factors affecting results in primary reading." *Twentieth Yearbook of this Society.* 1921, Part II, Chapter 1.

Zirbes, Laura. "Diagnostic measurement as a basis for procedure." *Elementary School Journal.* 18: March, 1918.

CHAPTER V
THE SIGNIFICANCE OF INTELLIGENCE TESTING IN THE ELEMENTARY SCHOOL

R. PINTNER
Teachers College, Columbia University, New York City

THE BEGINNINGS OF THE MENTAL TEST

The first mental test of any practical value for the measurement of intelligence was the Binet-Simon Scale. This scale was originally constructed to aid in the detection of feeble-minded children, and therefore, for a long time in the use of mental tests the emphasis was thrown upon the discovery of subnormal intelligence. It is from this period that we have inherited the expression "to submit a child to a mental examination," carrying with it a doubt as to the integrity of the child's intelligence. The need of society to protect itself against the feeble-minded was the reason for the development of the Binet-Simon Scale with its emphasis upon subnormal intelligence. If, for any reason, society had been more interested in the discovery of superior intelligence, the early history of mental testing would have been very different and it would have been regarded as more of a privilege than an indignity to be the subject of a mental examination. We have now, however, largely overcome the hostility and suspicion attaching to mental tests, and they are being used about as much for the discovery of superior intelligence as for the discovery of subnormal intelligence.

In addition to the individual examination, we now have the group examination, by means of which a large number of children may be tested at the same time. We shall, therefore, consider separately these two methods of examination and their value for the elementary school.

INDIVIDUAL TESTS

There are now many scales suitable for the individual examination of children. The ones most used at the present time are

the Stanford Revision of the Binet-Simon Scale,[1] the Goddard Revision of the Binet-Simon Scale, the Yerkes-Bridges Point Scale,[2] and the Pintner-Paterson Performance Scale.[3] The first three are revisions and extensions of the original Binet-Simon Scale, and of the three, the Stanford Revision by Terman is the best standardized and the one most extensively used. The Performance Scale makes use of none of the original Binet tests, but is composed entirely of form-boards and other performance tests, which do not require language either on the part of the examiner or the subject. It is therefore, extremely useful for testing foreign children; for children of foreign parentage where English is not spoken at home; for children suffering from speech defects of various kinds; for deaf children, and also as a supplement to any of the other scales which are so largely dependent upon language ability.

1. Service of Individual Tests in Locating the Backward

The main service which these individual scales render to the school at the present time is in the testing of children who are candidates for special classes of backward or bright children. Although group tests are being used to some extent for this purpose, it is generally felt that the more intensive individual examination is preferable. This is particularly true in the case of classes for the backward or feeble-minded, since unfortunately, a certain stigma sometimes attaches to relegation to such classes.

The segregation of subnormal children in special classes is now a firmly established policy in most progressive school systems. The selection of such children is generally, and should always be, based ultimately upon a mental examination. Because it is often impossible and unnecessary to give every child an individual mental examination, the usual policy is to ask the teacher to designate those children who are so poor in their school work as to arouse a suspicion of mental defect. These cases are then tested by the

---

[1] Terman, L. M. *The Measurement of Intelligence.* Houghton Mifflin, 1916.

[2] Yerkes, R. M., Bridges, J. W., and Hardwick, R. S. *A Point Scale for Measuring Mental Ability.* Warwick and York, 1915.

[3] Pintner, R., and Paterson, D. G. *A Scale of Performance Tests.* Appleton, 1917.

school psychologist or mental tester, and if they are found to be mentally inferior, they are then assigned to the special class. Children with an intelligence quotient below 80 should always, if possible, be given the benefit of instruction in special classes, and many children with I. Q.'s between 80 and 90 may profit by such special class work. There can, however, be no hard and fast line for the assignment of such children. The policy in each school system must depend upon the number and location of the available special rooms. Where the number of rooms is very small, it may only be possible to take care of the most retarded children. The special class may thus become filled with absolutely feeble-minded children, whose intelligence quotients are below 70. This is, of course, better than no segregation at all, but it does not take care of the borderline and backward cases with intelligence quotients ranging from 70 to 90, and a great many of these can profit by special class work. In some school systems a special building is assigned for the work with backward children, and this has the advantage of allowing a closer grading of the children, so that those of similar mental age may be grouped together. This grouping of children of like mental ability facilitates the work of the teacher immensely and is much more advantageous for the child.

It is needless here to attempt any survey of the progress of the special class movement in this country. Although in many respects much remains to be done, nevertheless, the growth of the work has been rapid and phenomenal, and it might not be an exaggeration to say that at the present time backward and feeble-minded children are receiving more attention and better instruction than any other group of children in our public schools. Most of this growth has been the result of the introduction of the mental examination, because the use of mental tests has clearly revealed the extent of the problem and has allowed us to make the selection of children accurately and quickly.

2. Service of Individual Tests in Locating the Superior

Only recently have we become definitely conscious of the presence in our schools of another group of children whose need for special instruction is as great as, if not greater than, that of the

backward and feeble-minded. The bright or superior child has been almost entirely neglected. He has been discovered by means of the mental test. After the first interest in the subnormal had subsided, it was inevitable that more and more attention should be paid to those children who were doing exceptionally well in the mental tests. The discovery of these cases was greatly facilitated by the appearance of the Stanford Revision of the Binet Scale, because this scale gave a much better opportunity than the original Binet Scale for a child to make a high mental age. Terman was one of the first to direct attention to the superior child and he has contributed a great deal to our knowledge of the subject.

Miss Race[4] at Louisville, Kentucky, seems to have been about the first to organize a special class for very bright children on the basis of mental tests. Whipple's[5] experiment in Illinois showed conclusively the necessity for the use of mental tests in the selection of children for such classes. It is well to emphasize this at the present time, because there is a tendency to believe that teachers and others are fairly well able to pick out the brightest children. This, however, is far from the truth. Most teachers are better able to select the mentally inferior than the mentally superior. If tests are useful for the selection of the dull and backward children, they are absolutely necessary for the selection of the mentally superior. A child who is doing the best school work in a class is not *ipso facto* a superior child. Superior intelligence and good school work do not always go together. There are many children doing only average or below average work, who are of superior intelligence. These children have simply formed the habit of doing passable school work, and they require a greater stimulus than the ordinary school provides to arouse them out of their apathy. Again many bright children are so bored by the slow pace of the average class that they lose all interest in school work and devote themselves enthusiastically to extra-school activities which give full play to their intelligence. The need of mental tests for a proper selection of such children is, therefore, obvious.

---

[4] Race, H. V. "A study of a class of children of superior intelligence." *J. of Ed. Psych.* 9: Feb. 1918, pp. 91-98.

[5] Whipple, G. M. *Classes for Gifted Children.* Public School Pub. Co., Bloomington, Ill., 1919.

Coy, at Columbus, Ohio, has conducted a very thorough and lengthy experiment with a special class of bright children. The members of this class were carefully selected on the basis of mental tests, and it was to this careful selection of cases that the success of the experiment was partly due. It was again demonstrated with reference to the selection of cases that dependence upon the choice of the teachers would have resulted in the omission of several of the very brightest and conversely in the inclusion of some of only average capacity. The homogeneity of intelligence in the group selected by the tests allowed the children in the class to advance together without the usual interference produced by the presence of slower and duller pupils. No attempt was made to set any definite pace in order to accomplish any given amount of the ordinary school curriculum. The children were allowed to set the pace and to cover as much as they seemed capable of doing, and at the same time, they were allowed to branch out into other subjects not generally included in the curriculum. Both enrichment of curriculum and acceleration took place. The question is often asked as to whether the curriculum ought to be broadened or whether it should be covered more rapidly. The question should not be stated in that way, as if these two things were mutually exclusive. In all probability, judging from Coy's work at Columbus, both enrichment and acceleration should occur in any carefully selected class of superior children. The class in question actually covered three years' work of the ordinary curriculum in two, and in addition received instruction in several subjects not found in that curriculum. When the class was abandoned, the children were ready for the eighth grade, and reports of their work in that grade show that they are doing much better than average work.

The experiment was eminently successful and revealed the great latent possibilities of the superior child. It aroused in them a desire to master things more difficult than they had ever met with before, and it thus gave them the opportunity of better gauging their own powers. Without some such stimulus as the special class provides, the great danger is that the superior child may go through life not dreaming of his potential ability, because school and society puts its approbation upon average work, and he may

have formed a habit in school of being content with this type of work.

This brief account of the selection of superior children must suffice here. Without doubt, the near future will see an increased interest in this type of special education. The number and variety of classes for bright children will unquestionably increase, when once we realize the big dividends they will pay. So far the interesting thing for the psychologist and the educator is to note the insistence of the pioneers in this work upon the necessity for mental examinations in the selection of the children. The Stanford-Binet has been most widely used. Group tests, as we shall see, are becoming increasingly valuable and accurate for classification purposes, but at the present time, wherever possible, a thorough individual examination is strongly to be recommended.

Group Tests

So far we have dealt with the use of individual scales and we have seen that the main use of such scales has been the selection of special cases, whether feeble-minded or superior. The individual examination is of necessity limited in scope in school testing, because of the amount of time necessary for the giving of a single test. There has, therefore, been developed within recent years the more economical group test, and its value to the school has exceeded the expectations of its most enthusiastic supporters. We shall discuss in this part of our article the chief group mental tests useful for the elementary schools and also the most important purposes for which they are being used. Tests for the first grades are described elsewhere in this *Yearbook*.

1. Some Group Tests Suitable for the Elementary School

The National Intelligence Tests. These tests were prepared under the auspices of the National Research Council by Haggerty, Terman, Thorndike, Whipple, and Yerkes. Two booklets are recommended for each examination. Each booklet contains five exercises.

Scale A contains (1) arithmetical problems, (2) sentence completion, (3) checking attributes possessed by a given word, (4)

synonym-antonym, (5) copying numbers corresponding to given symbols from a key.

Scale B contains (1) computation, (2) general information, (3) logical judgment, (4) analogies, (5) discrimination of similarity and difference as applied to numbers and forms.

The novel feature of this test is the fore-exercise that precedes each exercise proper. This fore-exercise is a sample of the kind of thing which is to be done in the test proper which follows immediately afterward, and thus gives the pupil an opportunity to adjust himself to the situation presented by the test. It is a preliminary practice period for each test, and the pupil's work during this period is not scored. In most cases the fore-exercise is limited to 30 seconds. Two forms of these tests have already been published, and three additional forms are promised. Each of these five forms will be equivalent to any other. Therefore, the tests may be used repeatedly without fear of coaching or of the pupils becoming too familiar with the specific questions of any one form. The tests have been given to thousands of pupils, so that good norms are available.[6]

The Haggerty Delta 2. This test is designed for grades three to nine. It is an adaptation of the Army Intelligence Examinations and was devised for, and used in, the Virginia School Survey. There are six exercises: (1) discrimination between true and false statements, (2) arithmetic, (3) picture completion, (4) discrimination between words, whether same or opposite, (5) common-sense judgments, (6) general information. This test is better adapted for elementary school purposes than the original Army Alpha. The norms consist of average scores for each age for ages eight to fifteen, and for each grade from three to nine. These average scores are based upon twenty thousand cases.

The Pressey Cross-Out Tests. These tests have been found useful in grades three to the high school. They differ from the tests previously described in that all of the four exercises call for the same type of response, namely, crossing out something; thus, Test 1, Cross out the superfluous word in disarranged sentences; Test 2,

---

[6] For a more detailed description of these tests, see Whipple, G. M. "The National Intelligence Tests." *Jour. of Educ. Research,* 4: June, 1921, pp. 16-31.

Cross out the superfluous word in lists of words related to each other; Test 3, Cross out the superfluous number in a number series; Test 4, Cross out the worst thing in several lists of qualities, actions, and the like. This last test is a sort of moral judgment test and differs radically from the type of test usually included in intelligence examinations. It seems to assume that a high degree of conformity with the conventional standards in moral judgment goes along with high general intelligence. Until we know more about such relationships, the test seems a little out of place in a general intelligence examination, but it is interesting in that it fore-shadows morality and character tests. There are excellent norms for these tests for ages ten to seventeen and for grades three to twelve.

The Otis Intelligence Scale, Advanced. This is suitable for grades five to twelve. It consists of ten exercises: (1) following directions, (2) opposites, (3) disarranged sentences, (4) matching proverbs, (5) arithmetic, (6) geometric figures, (7) analogies, (8) similarities, (9) narrative completion, (10) memory. This was one of the first tests to be published and it has been extensively used. The group tests used in the army were largely based upon the work of Otis. There are norms for ages eight to eighteen, inclusive.

There are several other scales which are useful in the upper grades of the elementary school and in the high school as well, for example: Terman's Group Test (grades 7 to 12); Dearborn's Scale II (grades 4 to 11); Whipple's Group Test (grades 4 to 8); Myer's Mental Measure (all grades); Pintner's Survey Tests (grades 3 to 10); Trabue's Mentimeters (all grades); and so forth.

2. The Use of Group Tests

The tests that we have mentioned have been more or less extensively used. Some are better constructed and better standardized than others. All of them will give a more or less accurate measure of a pupil's intelligence. It is impossible to answer the question so frequently asked: which is the best? The best for what purpose? Some of them are good for certain grades and have little discriminating power above and below specific limits. If extensive mental surveys of several schools or school systems are

to be made, several of the shorter tests will be found sufficiently accurate. On the other hand, where much depends upon the rating of the individual child, it is better to give the longer and more thorough tests, and still better to give more than one group test.

The one thing that any of these group tests will do is to rank any group of children in order of ability from the best to the poorest. This can be done regardless of whether there are good norms for the test or not, and this is after all the fundamental value of a mental test. The comparison of one pupil with another in reference to mental ability is the important thing, because the chief practical value is the grouping of children into more or less homogeneous groups with reference to their mental ability. The more alike in general ability the pupils in any one class are, the easier and more effective will be the teaching of that group. Now, one of the most striking results of the application of group tests to school children has been to show how very heterogeneous is the mentality of the children in an ordinary class. We find very superior, normal, backward, and dull children all grouped together and all expected to learn the same things and to learn them at the same rate. In the same class will be found children of quite varied mental ages.

One study[7] reports a range in mental age from four to nine in Grade I; from six to nine in Grade II; from six to twelve in Grade III; from six to fifteen in Grade IV; and similarly for the other grades. Terman[8] reports a range in mental age from three to ten in Grade I; from seven to fifteen in Grade V; and from twelve to nineteen in Grade IX. In a survey[9] of 1043 eighth-grade pupils in 29 schools in Oakland by means of the Otis Tests, it was found that the scores for the individual pupils ranged from 14 to 152 points, and that the medians for the 29 different schools ranged from a score of 48 to 109. As the examiners point out, the mental ability of the best eighth grades was as good as that of an average ninth grade, and the mental ability of the lowest eighth grades

---

[7] Pintner, R., and Noble, H. "The classification of school children according to mental age." *Jour. of Educ. Research.* 2: Nov. 1920, pp. 713-728.

[8] Terman, L. M. "The use of intelligence tests in the grading of school children." *Jour. of Educ. Research.* 1: Jan. 1920, pp. 20-32.

[9] Dickson, V. E., and Norton, J. K. "The Otis Group Intelligence Scale applied to the elementary school graduating classes of Oakland, California." *Jour. of Educ. Research.* 3: Feb. 1921, pp. 106-115.

equalled only that of an average sixth grade. Colvin[10] reports pupils in Grade VII ranging in score from 27 to 143 points on the Otis Scale; in Grade VIII ranging from 47 to 171 points.

Such results are typical of what has been found in every school survey by means of mental tests. We are slowly coming to a realization of the tremendous differences in mentality that exist in children of the same chronological age. To some extent this is already beginning to affect school procedure in the grouping of children, although for the most part we are still under the incubus of chronological age. In course of time, however, when the significance of the results of mental tests becomes more widespread, we shall gradually pay less and less attention to chronological age and more and more to mental age.

The Combination of Mental and Educational Tests

It is obvious that these radical diffierences in mental ability among children of the same class, .among children in different classes, among different schools and school systems, affect very materially the amount of educational attainment achieved by various groups. A child of inferior mentality cannot be expected to accomplish educationally as much as a child of superior mentality. In the same way, a class or school with a low average mental ability should not be expected to cover the same curriculum as quickly as a class or school with a higher mental ability. The relationship between mental ability and school progress in the individual child has for a long time been recognized, and opportunities for slower or faster progress have been allowed for by the formation of special classes, as we have already noted. The fact that there are appreciable differences in mental ability among ordinary classes and schools is only now being slowly recognized. Up to the present time it has been tacitly assumed that the average ability of any class or school was equal to that of any other class or school and that, therefore, it was reasonable to expect the same amount of educational progress in each case. All grades in a school system are expected to cover the same amount of the course of study laid

[10] Colvin, S. S. "Some recent results obtained from the Otis Group Intelligence Scale." *Jour. of Educ. Research.* 3: Jan. 1921, pp. 1-12.

down for the system, making no allowance for the different mental abilities of the classes or schools. If one school falls below another in educational achievement, it is generally assumed to be the fault of the teachers and principal of the school. The fact that there are great differences in the raw material with which teachers have to work has seldom been fully recognized. The raw material with which the teacher has to work is the native ability of the child, and this determines the degree of modifiability or the rate of learning. Good raw material is easily modifiable and the rate of learning is rapid. Poor raw material is hard to modify and the rate of learning is slow. A teacher should not be blamed for the poor raw material with which she may have to deal. But, we should see to it also that she makes efficient use of the good raw material.

A serious defect of most school surveys up to the present time is the lack of a measure of the intelligence of the pupil material. The best of these surveys have made excellent use of objective educational tests and scales, and the results have been of great value. Many of the conclusions drawn from these results are, however, open to criticism. If a school or class is below the average in any given subject, the suggestion has been that the administration of the school, the attendance of the pupils, the physical equipment of the school, and particularly the methods and teaching ability of the staff are at fault, and it has been upon the teachers that for the most part the blame has rested. Now, poor teaching will undoubtedly lead to slow educational progress, but from the results of combined educational-mental tests that we are now getting, we have reason to believe that poor teaching is more likely to be found in schools possessing good mental material than those possessing poor mental material, because in the latter there is constant pressure being brought to bear upon the teacher to cover the regular course of study made out for the school system as a whole. The basic differences in the mental ability of the pupils, which in all probability are the chief reason for the differences in educational attainment, are seldom mentioned or when mentioned, seem to be considered of secondary importance.

Survey Results

The Cleveland Survey[11] gives excellent tables and diagrams showing the differences that exist among schools in various educational subjects measured by standard tests. Thus, in one arithmetic test the median score in the eighth grade for 90 schools is 27.5, but the range of medians is from 21 to 41. The same wide range appears in the other grades. In reading, in the fourth grade the scores for 44 schools range from 34 to 63, with an average score of 47. The other school subjects measured show similar enormous variations from grade to grade.

In attempting to interpret these differences the survey report never emphasizes the differences in the mentality of the pupil material. In fact, this is scarcely ever mentioned. To be sure, the report says that "children in different schools differ from one another," but it does not go on to explain what kind of differences are meant, and one gets the impression, because of frequent mention, that differences in nationality and social condition are the differences considered important. Again, the report says that "it becomes necessary at times in reporting the results of the tests to criticize the schools which are below the average, or are irregular in their instruction," from which teachers and principals draw the natural conclusion that if their schools are below the average, they themselves are more or less to blame. In many cases the educational work in schools below average is as good as we have a right to expect in view of the ability of the pupil material. Again, the report continues: "Every adverse criticism based on comparison thus implies praise of the good school and the excellent work which furnished the basis of comparison." This, of course, implies that work above the average is due to the efficiency of the teachers and principals, whereas, as a matter of fact, we have reason to believe that it may be solely due to the mental make-up of the pupil material, and in many cases such educational work is not nearly as good as it ought to be in view of the excellent native ability possessed by the pupils. Praise or blame, therefore, cannot be apportioned on the basis of educational tests alone. To judge justly of the

[11] Judd, C. H. *Measuring the Work of the Public Schools.* Survey Committee of the Cleveland Foundation, 1916.

work of a school, we must have a measure of the mental ability of the children.

We have taken the Cleveland Survey as a sample of the best type of recent school surveys, and we do not mean to suggest that the writer of the report was not aware of differences in mentality in different schools. In many other surveys the neglect of such differences is much more flagrant. In all surveys up to the present time, the great amount and the importance of such differences have not been fully realized.

Combined Measures

Several workers have pointed out the necessity for an evaluation of educational attainment in terms of mental ability. The writer[12] suggested this in 1918 and in more detail in 1919.[13] In 1920 Franzen[14] proposed the A. Q. or Accomplishment Quotient. The A. Q. is the E. Q. (educational quotient) divided by the I. Q. (intelligence quotient). The I. Q. is a measure of the native ability of the child and shows his potential rate of progress. The E. Q. is a measure of the educational attainment of the child and shows his actual rate of progress. "The Accomplishment Quotient is the degree to which his actual progress has attained to his potential progress by the best possible measures of both." And further: "It is a mark which evaluates the accomplishment of the child in terms of his own ability. A brilliant child would no longer be praised for work which in terms of his own effort is 70 percent perfect, in terms of the group, 90 percent . . . A stupid child who does work which is marked 70 in terms of the class, but 90 in terms of his own, a limited ability, is no longer discouraged."

Two sets of tests have been recently published for obtaining a combined educational-mental measure, although, of course, an E. Q. and A. Q. as suggested by Franzen can be obtained wherever we have mental and educational tests standardized by ages. The

[12] Pintner, R. *The Mental Survey.* Appleton, New York City, 1918.

[13] Pintner, R. Paper read before the American Psychological Association, Dec. 1919. *Psychol. Bulletin.* 17: Feb. 1920, pp. 60-61.

[14] Franzen, R. "The accomplishment quotient." *Teachers College Record.* 21: 1920, pp. 432-442.

writer's[15] combined mental-educational tests have been specifically devised and standardized for general survey purposes to give a rough measure of the intelligence and the educational attainment of pupils in the elementary school from Grades III to VIII. The Illinois examination by Buckingham and Monroe[16] contains a mental test of seven exercises, and two educational tests, namely, reading and arithmetic, and is suitable for Grades III to VIII.

We have thus seen in a relatively short time the principle of evaluation of educational attainment in terms of mental ability very definitely stated, various means for such an interpretation suggested, and two combined sets of tests published. Let us now look at some of the more striking results that seem to be emerging.

The thing that has impressed the writer most in his own work is the seemingly greater inefficiency of the brighter children, when they are measured with reference to their potential ability. Thus, in tests of 4215 children, of the 900 children doing less than their mental capacity would seem to warrant, 47 percent are diagnosed as bright by means of the intelligence test and only 8 percent as backward. Again, of the 1064 children who seem to be doing more than is generally done by children of like mentality, only 11 percent are bright mentally, while 40 percent are mentally slow. The results obtained may be seen in the following table:

	Doing less than expectation	Working up to expectation	Doing more than usually accomplished
Bright	47.4	24.4	10.8
Normal	44.3	53.2	49.3
Backward	8.3	22.3	39.8

It is evident, therefore, that the tendency of the school is to push ahead the mentally slow in order to make them keep pace with the average and at the same time to neglect the bright as soon as they have achieved average work.

---

[15] See Pintner, R. *Manual of Directions for Combined Mental-Educational Tests.* College Book Co., Columbus, O.; and also Pintner, R., and Marshall, H., "A combined mental-educational survey." *Jour. of Educ. Psych.* 12: Jan. 1921, pp. 32-43, and 12: Feb. 1921, pp. 82-91.

[16] Buckingham, B. R., and Monroe, W. S. "A testing program for elementary schools." *Jour. of Educ. Research.* 2: Sept. 1920, pp. 521-532.

What is true of the individual child seems also to be true of the school in general. We find many schools where the general ability of the pupil material is excellent, that are failing to live up to their possibilities in the way of larger educational returns; and, conversely, we find many schools of poor pupil material that are giving relatively good educational returns, even though the absolute accomplishment seems poor. We cannot, therefore, justly evaluate educational accomplishment without some measure of the ability of the pupil material. Although most of these results at present point to a tremendous wastage of good intelligence, we may be optimistic as to the future when we hope that this intelligence will be discovered early and be thoroughly utilized.

Summary

We have attempted to show in general the place of mental testing in the school, both from the standpoint of the teacher and superintendent, as follows:

1. The use of individual tests as a means of careful diagnosis, where special educational treatment of specific pupils is concerned.
2. Individual tests useful for the selection of dull and bright children in the organization of special classes.
3. The use of the group test for the classification of children so as to group together children of like mentality.
4. The various kinds of group tests at present available for the elementary school.
5. The need of both educational and mental tests in the evaluation of the work of the teacher and the principal.
6. Various measures proposed for such evaluation.
7. Some consequences of the use of such combined mental educational measures.

CHAPTER VI

THE USE OF INTELLIGENCE TESTS IN JUNIOR HIGH SCHOOLS

M. R. Trabue
Assistant Professor of Education and Director of the Bureau of Educational Service, Teachers College, Columbia University, New York City

Only in so far as the junior high school differs from other segments of the educational establishment will the uses of intelligence tests differ in a junior high school from their uses in other schools. The most outstanding characteristic of the junior high school is undoubtedly its sensitiveness to individual differences in pupils. This responsiveness to differences in its pupils is largely the result of fundamental purposes, although partly an accident due to the newness of this type of school. Furthermore, unless attention to differences is fostered and held constantly in mind as a cardinal virtue, such a school will soon lose the majority of its distinctive features.

If one takes the five peculiar functions of the junior high school found by Koos[1] to be mentioned most frequently in school documents and in the statements of educational leaders about such schools, he may recognize each function as being to a large extent a result or an expression of the responsiveness of the junior high school to the differences existing in its individual pupils. These five functions are:

I. Realizing a Democratic School System through
   A. Retention of Pupils
   B. Economy of Time
   C. Recognition of Individual Differences
   D. Exploration for Guidance
   E. Vocational Education
II. Recognizing the Nature of the Child
III. Providing Conditions for Better Teaching
IV. Securing Better Scholarship
V. Improving the Disciplinary Situation and Socializing Opportunities

[1] L. V. Koos, *The Junior High School* (New York: Harcourt, Brace and Howe, 1920), p. 18.

Pupils are to be retained in larger numbers by the junior high school, because it recognizes that they are not all interested in the same kind of work and therefore provides a greater variety of courses than the usual grammar school, with some opportunity for the individual pupil to choose what he will study. Time is to be economized in the junior high school by recognizing that some of the traditional subject matter is of little value to most of the pupils and by grouping pupils according to their abilities to make progress. Certain courses are to be given primarily as introductions to the essential facts and skills in different types of trades and occupations from which each pupil may later choose the one in which he may find his greatest interest and probable success. Better teaching, better scholarship, better discipline, and better social organization are to be secured through the grouping together for study and recitation of pupils who have approximately the same abilities, and through the recognition by the school and exercise by the pupils of different degrees of social, political, and administrative powers.

Obviously, the most important use of intelligence tests in the junior high school will be the discovery and measurement of differences in the intellectual abilities of the individual pupils. Although desirable traits tend to be found in the same individuals, the correlations between intelligence and such qualities as moral honesty, industry, social leadership, and political sagacity are not perfect. It will not be possible, therefore, to measure by means of intelligence tests all of the individual differences to which the junior high school must give recognition and make adjustments. In so far, however, as the type of intelligence measured by our tests is the type to which the school should be sensitive, intelligence tests are indispensable tools in the organization and administration of the modern junior high school.

If it were possible to measure with great accuracy every type of capacity and ability, no two pupils would be found to be alike. Each individual pupil probably has a different degree of native intellectual power, a different amount of social instinct, a different quantity of self-control, and a different avoirdupois weight from any other pupil in the same school, although our scales for measuring

these qualities are sometimes so crude that we can not distinguish the differences. As a matter of fact, although such differences do exist, they are frequently so small as to be of no vital importance so far as life or the school is concerned.

Considering the matter abstractly, a thoroughly democratic state should provide each child an equal opportunity to develop his individual capacities to their maximal effectiveness. To ignore the fact that children differ in their native endowments and in their social and vocational futures, and to force all pupils to take exactly the same educational course is not only extremely undemocratic, but is also practically impossible. However narrow and uniform the offerings of a school may be, its pupils do not obtain the same amounts of training from the same amounts of attendance. If individual differences in children were the only factors to be considered in the formulation of an educational program, individual instruction would be the universal practice, not only in regard to the rates of progress, but also in regard to the fields in which progress would be attempted.

From an economic and social point of view, however, it would be extremely wasteful of the energy of teachers and of the public resources to train each child separately. A public school must serve the state economically as well as serve the future citizens of the state individually. Certain differences in children's endowments and future histories are so small as to be relatively unimportant as far as their training in a given field is concerned. Furthermore, there are certain habits of thought, action, and feeling which must be more or less universal if the state is to maintain itself as a unit. For these and other reasons, pupils in the public schools are grouped in classes, rather than taught as though each individual were a distinct class in himself.

It was stated above that the junior high school is characterized by its unusual sensitiveness to individual differences. Being less closely bound by tradition than other schools, the degree to which the junior high school may adjust itself to differences in its pupils is controlled chiefly by economic and social expediency. The size of classes must be such as will give the maximal opportunity to each individual pupil without the expenditure of more time, energy,

and money than the general public can approve and supply. The variety of subjects offered must meet as far as possible the individual needs of all the pupils, but must not be so great as to take, for the training of a few, public funds which are more definitely needed for the instruction of many. Although "an attempt to provide differentiation is the most marked characteristic of junior high schools,"[2] the extent to which this attempt may be carried is limited by the size and wealth of the community and by many other factors.

Such studies as have been made of measured differences in the intellectual abilities of secondary school pupils indicate two uses to which the results of intelligence tests may reasonably be applied in the differentiation of junior high school pupils. The results obtained from intelligence tests now available may be used as one element in the prognostication of the field of the pupil's probable educational and vocational future, pointing out for him the program of studies and work which will be of greatest usefulness to him; and they may be used in the prediction of the rapidity with which the pupil will be able to make progress in his studies. In other words, the results of intelligence tests may be used as one means of helping a pupil choose wisely the direction in which he should go, and then as a means of so classifying him that he will be associated with others who are going not only in the same direction but also at the same rate.

Most of the evidence that intelligence tests may be used as a basis for the guidance of pupils into the educational or the vocational field where they would be most successful, has been obtained by measuring the intelligence of pupils who of their own choice have already entered upon certain educational or vocational careers. The argument, therefore, is seldom that pupils divided and assigned on the basis of these tests were successful in certain courses or trades, but more frequently that pupils who made choice of these lines of work and were then successful in them, made such and such scores when measured by the tests; and therefore

[2] Briggs: *The Junior High School* (Houghton Mifflin Company, 1920), p. 154.

that those who make such and such scores would undoubtedly be successful in these lines of work or study.

Determining the coefficient of correlation between the tests of intelligence and the school success of the pupils has been a popular method of determining the usefulness of intelligence tests in the guidance of pupils and was the method used by Wood at Kansas City, Mo., with a first-year algebra class in 1917.[3] The Stanford-Binet Tests of Intelligence and the Rugg and Clark Algebra Tests were given in a first-year algebra class. The coefficient (by the Spearman Foot-Rule) between intelligence quotients and class grades was .993, while the coefficient between the arithmetic means of all marks in the sixteen Rugg-Clark tests and the intelligence quotients was .998. Such unusually high correlations would not often be obtained, especially if computation were by the standard product-moment method (Pearson-Brevais), but the report is of interest. "Since there is a close relation between general intelligence and ability to learn algebra, it seems reasonable to conclude that the general intelligence of each pupil should be determined before he is required to take the subject. If he is clearly below normal in general intelligence, he should be prohibited from taking algebra unless there should be good reasons to the contrary."

Madsen reported the relationship of the Army Alpha Tests to success in the high schools of Omaha, showing that a difference of 20 to 30 points existed between the scores of corresponding classes in the Central High School and in the Commerce High School.[4] The differences in the scores obtained by pupils studying different subjects were so marked that Madsen concluded that "either the standards for success are relatively lower for the vocational subjects taught in Commerce High or a less degree of intelligence is required for success in them."

One of the most careful workers in this field is Professor Proctor of Leland Stanford University. During the school year 1916-1917 he examined 107 high-school pupils by means of the Stanford-Binet

---

[3] O. A. Wood: "A failure class in algebra." *School Review,* 28: pp. 41-49.

[4] Madsen, I. N. "Group intelligence tests as a means of prognosis in high school," *Journal of Educational Research,* 3:43-52; and "Relationship between general intelligence and success in certain high-school subjects," *Journal of Educational Research,* 3:396-398.

Scale and compared the results with the school marks earned during that year and with the teachers' estimates of intelligence.[5] Two and a half years later only 66 of the original 107 remained in the same high school; 20 of them had transferred to other high schools, and 21 had left school to go to work.[6] The average school rating of those who went to work was 73; of those who transferred, 77; and of those who remained in the same school, 79. The median intelligence quotient of those who went to work was 94, that of those who remained in school was 110. Of those who were originally found to have I. Q.'s below 90, only 25 percent remained in school at the end of a year, while of those having I. Q.'s above 110 it was found that 100 percent were still in school at the end of two and a half years. The correlations of the intelligence quotients of the 107 pupils with teachers' estimates of intelligence was .586, ± .043, and that with the average of school marks was .545, ± .046.

Similar study of the records of 955 high-school pupils tested in 1917-1918 by the Army Intelligence Tests, showed two years later that of those remaining in the high school only one-fourth had I. Q.'s below 100, while of those who had gone to work more than 60 percent had I. Q.'s below 100. As the result of these findings, Proctor believes that "discovering at the outset that from 15 to 30 percent of his (the principal's) pupils are incapable of succeeding in the conventional high-school subjects, he will undertake to make new adjustments to meet the situation. There will be fewer failures; more pupils will remain to take work that is adapted to their needs and capacities; and the high school will be less open to the charge of catering only to the intellectual aristocracy among its pupils."

Proctor has also furnished the most definite report showing the actual success of educational guidance.[7] This report gave measures of the relative success of two groups of pupils entering the high

[5] Proctor, W. M. "The use of intelligence tests in the educational guidance of high-school pupils," *School and Society*, 8: pp. 473-478, 502-509.

[6] Proctor, W. M. "Psychological tests as a means of measuring the probable school success of high-school pupils," *Journal of Educational Research*, 1: pp. 258-270.

[7] William M. Proctor: *Psychological Tests and Guidance of High School Pupils*. (Bloomington, Ill.: Public School Publishing Co., 1921.)

school, one group having been carefully advised individually as to the work that should be undertaken and the other group having made their own selections of courses in the usual manner, although both groups had been examined by means of intelligence tests and found to be equally capable. That success in the first year of the high-school course is more certainly assured to the pupils who are guided in the selection of their courses is clearly indicated by the following table, adapted from page 30 of Proctor's report.

SUCCESS RECORDS OF FIRST-YEAR HIGH-SCHOOL PUPILS WHO WERE "GUIDED," COMPARED WITH THOSE "NOT GUIDED"

Group	No. of Pupils	Percent Left to go to Work	Percent Transferred to Other H. S.	Percent Failed in	
				One Subject	Two Subjects
Guided	22	4.5	9.1	18.2	0.0
Not guided	107	12.1	13.1	30.8	10.3

The evidence in favor of vocational guidance in the junior high school is less abundant and direct than that in favor of educational guidance. The argument is again that those who belong to a certain group of trades or vocations make scores of a given size, and therefore that pupils who make scores of a given size may expect success in a given group of vocations, provided they have the other qualities and training needed to supplement their intellectual gifts.

The most extensive study bearing on this subject was conducted by the Division of Psychology of the Office of the Surgeon General, U. S. Army in 1918.[8] The intelligence test records of soldiers who claimed to belong to various occupational groups were studied, with results which may be of some value in the vocational guidance of pupils in the junior high school. Only selected vocations are given in the following table, and the grouping is that of the present writer rather than of the Division of Psychology. The table gives the average or median score of each vocational group of soldiers on Test Alpha, with the range of scores necessary to include the middle half of all scores made by the group.

[8] *Army Mental Tests: Methods, Typical Results and Practical Applications* (Washington: Government Printing Office, 1918). See also C. S. Yoakum and R. M. Yerkes, *Army Mental Tests* (Henry Holt and Co., New York, 1920), especially pp. 196-203.

TYPICAL SCORES FOR OCCUPATIONAL GROUPS IN THE ARMY. INTELLIGENCE TEST ALPHA

Occupations	Score Median	Interquartile Range
Workers with simple tools and materials	—	21–83
Laborers	35	21–63
Teamsters	41	23–68
Farm Laborers	42	24–70
Horse-shoers	44	25–70
Bricklayers	48	23–81
Painters	53	31–79
Blacksmiths	54	29–83
Workers requiring considerable skill	—	33–99
Carpenters	57	33–85
Butchers	58	33–85
Machinists	61	33–86
Plumbers	62	38–87
Chauffeurs	63	38–90
Telephone operators	70	58–99
Workers requiring high-grade skill and knowledge	—	52–133
Photographers	77	52–104
Electricians	82	58–110
Telegraphers	84	59–107
Mechanical engineers	98	63–133
Workers with symbols and ideas	—	78–
Bookkeepers	99	78–126
Stenographers	115	93–142
Accountants	117	101–145
Civil engineers	125	98–147
Physicians	130	101–165

Although the studies just mentioned and many others of a similar nature indicate the probability that an intelligence test score of a certain size may be used as a fairly good index of the vocations or courses of study in which the child might expect success, the public in general will wish to have further evidence from the actual success or failure of children who have been guided into the vocations or into the educational courses on the basis of the results of intelligence tests. Furthermore, it is quite clear that one can not use the test results alone as a basis for the guidance of pupils, for a given score in such a test may be typical of successful persons in a half dozen or more different specific vocations or curricula. The interpretation of the intelligence tests in educational and vocational guidance is largely negative, suggesting lines of work in which the child will probably fail rather than asserting that the individual

will be successful in a given field. Tests of aptitude and probable success in specific lines of endeavor are much needed by those engaged in guiding young people. Such specific tests, used in the junior high school in connection with courses for the exploration and discovery of vocational interests, would supplement the negative evidence of the intelligence tests and make a real science of vocational and educational guidance.

Objection has arisen in some quarters to the idea of advising pupils as to their futures on the basis of scores in tests. The claim is made that such a procedure is undemocratic and that it closes the door of opportunity to many who might otherwise enter the "higher walks of life." It is asserted that if a pupil is placed in "practical" courses at the junior-high-school age, he is being condemned to a "level of activity" which may not be the highest of which he is capable. The argument is usually that the pupil should be allowed to continue taking the general or academic course until he reaches a place where he can not make further progress, and then as a last resort he may be given some vocational instruction, *provided he has remained in school.*

If a pupil once started on a semi-vocational course is to be refused permission to return to an academic course, or if the advisor uses autocratic power and insufficient evidence, placing pupils mechanically according to their test scores and without regard to the pupil's interests and to other obtainable criteria, then certainly no right-minded person would argue for such vocational guidance in the junior high school. The tests at present available are so inadequate and crude that one who uses a single test score as the sole basis for a vital decision in the life of an American youth is guilty of most unscientific practice and possibly of a great injury to the child advised. Those who undertake to give educational or vocational guidance either in the junior high school or in more advanced grades must be persons of broad outlook on life, with a mature, well-balanced fund of active common sense and a clear understanding of the reliability and validity of the tests they employ.

Measurements of differences in the intellectual abilities of junior-high-school pupils, when supplemented by measurements of

their educational achievements and by the judgments of their teachers, may nevertheless be given most serious consideration in planning for the educational or vocational futures of boys and girls. Informing a pupil on the basis of such evidence that it would probably be useless for him to attempt to prepare for law, the ministry, or the "learned professions" might cause a momentary disappointment, but it would be less keen and less humiliating than the frequent failures in his studies and the constant struggle of working at tasks beyond his ability which would be certain to result from ignoring such predictions. Pupils guided by such evidences are not "condemned." They are rather "freed" from the prospect of being "failures" in school and probably even after they have left school. It is the pupils who are not given the opportunity in school to work at tasks which interest them and are not too difficult for them who are "condemned." The "single-track school" forces a large proportion of its pupils into the habit of expecting and achieving failure, which is certainly wrong from a moral and social point of view as well as from the personal standpoint of the one who fails.

Another misconception, implied in the opposition to the guidance of pupils, is that it is more noble and worthy for a pupil to take an academic course leading to the professions than it is to take a course leading to a trade. The maximal success of the world depends upon having each person do as well as he can the work for which he is best suited. The blind man does not feel that he is disgraced because he is not made an engineer on a railroad, nor does the man without musical talent condemn the world for not encouraging him to be a grand-opera singer. In a similar manner, those who are not gifted in the handling of ideas and symbols should not resent it if they are discouraged from becoming preachers and mathematicians, and those who have no interest or ability in mechanics should not chafe at being warned away from engineering as a profession.

Teachers are possibly to blame for some of the tendency to speak of the ability of the professional man as "higher" than the ability of the mechanic or laborer. Ability to use ideas, words, and symbols is not "higher" but is "different" from the ability to use

tools and raw materials. Both types of ability are necessary and entirely respectable if used for the common good. Measured by the scale of the laborer's ability, teachers would usually test "lower" than laborers, while on the scale of ability as a teacher one would no doubt find the teachers "higher" than the laborers. Teachers must take a broader view of the various life activities and realize that it is just as "high" and respectable to be a good street sweeper as it is to be a good teacher or lawyer. If the junior high school is to be a democratic institution, it will attempt to discover the differences in pupils' special gifts, and to train each pupil to be happy and effective in making his particular contribution to human happiness as efficiently as possible.

Intelligence tests are useful, not only in the educational and vocational guidance of junior-high-school pupils, but also in the grouping of such pupils for recitation purposes. Dividing an entering class into recitation sections according to the alphabetical list of names of the pupils is usually more satisfactory than dividing them according to the seats they happen to take on the first day of school, because the alphabetic scheme tends more certainly to secure groups of approximately the same *average* abilities. Within each group selected on the basis of the alphabet, however, a great range of educational and intellectual ability will be found. The slow, average, and rapid pupils will be associated together in each class. It is an economy of time for all concerned to have each recitation section composed of pupils all of whom have approximately the same degree of ability to make progress. Those who have tried them assert that the results of intelligence tests are an excellent partial basis for making up such homogeneous groups.

One of the earliest attempts at homogeneous grouping of junior-high-school pupils was that made under the supervision of Professor Thomas H. Briggs,[9] in 1915, at the opening of the Speyer experimental junior high school, which is operated jointly by the City of New York and Teachers College. The elementary school

[9] For a full report on this experiment see the article by Dr. Briggs in the *Third Yearbook, National Association of Secondary School Principals* (Menasha: George Banta Publishing Company, 1920), pp. 53-62, entitled "Provisions for Abilities by Means of Homogeneous Groupings."

marks for the 275 boys who were entering this school from the sixth grades of five or more public elementary schools and the score of each boy in each of ten psychological and educational tests were secured. Extracts from Briggs' report follow:

"On the basis of these records the boys were ranked according to estimated ability and divided into groups of twenty-five, the limit being set by the number of seats in the recitation rooms. In the first weekly conference the teachers were informed of this phase of the experiment and told that the grouping was tentative, to be modified whenever they could agree that any two boys should change places. They were told, too, that they were expected to carry each group forward at a speed that seemed best for its powers of learning.

"At the beginning of four successive terms new groups of pupils who entered the school were similarly classified, each having been measured with new combinations of tests, the effort being to secure a battery that could be taken by a considerable number of pupils simultaneously and that could be scored with the most economy of time and effort. . . .

"As the term progressed the teachers from time to time made transfers of pupils from one section to another, usually because it became apparent that they had been badly classified. In a number of cases, however, the transfer was reversed a few weeks later and the pupil found himself in the same group as before. . . .

"At the end of each term, the teachers were requested to rank in the order of ability all of the pupils in their classes. From these rankings, which were entirely separate from the marks given for class achievement, was made a composite ranking to represent the best judgment of the entire corps as to each pupil's relative ability, whether he exercised it consistently on his lessons or not. That even this composite ranking was inaccurate goes without saying. . . . On the whole, the teachers agreed very well among themselves in their estimates of pupils' general ability, but a study of their reports leads to the conclusion that a group of representative public school teachers, all interested in their work and with their attention constantly directed toward the pupils as individuals, are, after months of instruction in classes of ideal size, unable to judge with anything like accuracy the relative ability of their pupils. . . .

"Both the prognosis made from earlier school marks and that from the standard tests proved highly significant of what the pupils would do in their subsequent work. In the order of their merit, we found a composite of all sixth-grade marks least indicative of what the boys would do, a composite of all marks in Grades I to VI, inclusive, somewhat better, and the ranking by the tests easily best of all.[10] In fact, if I had to rely on the rank given a boy

[10] For the details of this study of the various means of predicting success, see Fretwell: *A Study in Educational Prognosis* (New York: Teachers' College Contributions to Education, No. 99, 1919).

after two hours of testing or on the judgment of the average teacher who had him in class for five months, I should with little hesitation choose the results of the tests. But even the previous school record, especially when supplemented by the grade teacher's judgment, will assuredly afford a classification better than that based on the alphabet, the neighborhood, or chance selection. Let me repeat again that any such classification as has been proposed should be only tentative, to be modified whenever it appears that a pupil can work to better advantage in another group.

"If the plan of homogeneous grouping is to prove successful, the teachers must be closely supervised, especially in the first few months. Being accustomed to attempt the same amount with each section of a class, the average teacher finds it difficult to break sharply from the practice. . . . The teachers must be led to find what the optimum pace for each group is and supervised until they learn to maintain it. In conference the teachers and principal should at the beginning of the term estimate approximately what each class may be expected to do, and then, as under the plan now in general use, progress should be roughly regulated by the program. . . .

"The ideal is to segregate pupils as homogeneously as possible and then to advance each group at its optimum pace, whether that be half normal or three-fourth normal or one and one-fifteenth normal. Any difference that results in substantial progress of the group without the unnecessary retardation of some and the discouraging failure of others equally earnest is surely worth seeking. . . .

"In no single instance have we felt that a pupil lost anything material by his classification; in the great majority of cases, the pupils were happier in their work and made better progress than they otherwise could have done. Some saved a year in their secondary school education, some a half-year, and some nothing at all; but none who remained a full two years (the elimination was very small) failed to be certified by their teachers as satisfactorily doing a full two years' work. Gratifying results have been manifest in the teachers themselves: their work has been more interesting, they have had less strain, and they have felt better satisfied with the results than under the usual organization. All of them have testified that they never wish to return to a plan whereby the classification is fortuitous and the expected progress uniform."

An interesting attempt at homogeneous grouping of pupils in the Washington Junior High School, Rochester, New York, has been reported by Glass.[11] Pupils entering this school in September, 1919, were classified, on the basis of their results on the Otis Group Intelligence Tests, the Terman Vocabulary, and the Chicago Rea-

[11] J. M. Glass: "Classification of pupils in ability groups," *School Review*, 28: pp. 495-508.

soning Tests, into full-schedule classes, three-fourths-schedule classes, and study-coach classes, the last being the pupils of the lowest scores in the intelligence tests. Teachers were not informed of the relative ranks of the groups, but through their contacts with the groups each teacher soon came to understand correctly what the ranks were. A repetition of the tests in February, 1920, gave the groups the same ranks, although individual pupils were somewhat changed in scores and in ranks.

Glass seems to feel a considerable degree of confidence in the tests as rough sieves for the first classification of pupils in the junior high school, but finds them inadequate for fine distinctions. Although justice seems done to each group, he finds that there is individual injustice in a few cases. He agrees with Briggs in urging the importance of the reclassification of individual pupils whenever later evidence from additional tests, teachers' experiences or retesting seems to warrant it.

Superintendent Callihan tried an experiment in which he employed the results of the Illinois Examination as one element in classifying the eighth-grade pupils at Galesburg, Illinois.[12] The tests were given in May, 1920, to all seventh-grade pupils who were going into the eighth grade. Mr. Callihan reported as follows:

> "The scores were tabulated and the pupils from all the seventh-grade rooms in the city were classified on the basis of these results and placed in homogeneous groups. Eight rooms were available in a central building, and here the two hundred and eighty-five eighth-grade pupils were brought together. For the sake of clearness the rooms were lettered A, B, C, D, E, F, G, and H. The students ranking lowest in intelligence were placed in Room G; the next in Room H, and so on up the scale to Room B. In Room A those pupils were placed who had already been in the eighth grade one semester and whose I. Q.'s were approximately normal. The lowest group was placed in Room G rather than in Room H, so that the designating letter would not indicate to the pupils whether they were in the best or the poorest room.
>
> "A course of study was then worked out for each room. For example, we expect the pupils in Room G to do only the minimum essentials for promotion; Room H does all that Room G is required to do, plus an additional amount; Room F is required to do still more; and so on up the scale until

[12] T. W. Callihan: "An experiment in the use of intelligence tests as a basis for proper grouping and promotions in the eighth grade." *The Elementary School Journal*, 21: pp. 465-469.

Room B is reached. In this room those pupils whose I. Q.'s ran above 120 were placed, and they are permitted to advance through the regular course of this grade as rapidly as they are able.

"When school opened in September, pupils in all the rooms except A and B were given to understand that they might be advanced to a higher room provided their work was above the average for their room. It was also explained that if they did not keep up with the others in the room, they would be demoted to a lower room. It has been necessary thus far to make only five transfers, three of which were promotions and two were demotions, a fact which is very good evidence of the reliability of intelligence tests as a means of grouping pupils on the basis of ability.

"In order to check up the results of the test given in May, 1920, the same test was given in October, 1920, the results placing the rooms in exactly the same order as they were placed by the first test.

"Up to the time that this article was written, Room B had completed a little more than half of the regular work of the complete eighth-grade requirements, and the semester was not then half over. In fact, in some lines the pupils were far ahead of the pupils in Room A who had spent one-half year in the eighth grade before entering in September. . . . If the pupils of Room B continue to progress as we believe they will, they should complete the last five years of their elementary and secondary school work in at most four years. In doing this, instead of forming habits of indolence and 'get by,' they will form habits of industry and 'do your best' which will carry over into their work which is to follow."

The most fundamental objection to the classification of pupils into groups of homogeneous intellectual ability is that such a group would lack certain differences between individuals which will almost certainly characterize every other group in which the pupil may later live. The argument is that the bright pupil would not have the opportunity to develop his capacity for leadership in a group of pupils as bright as he, at least not as great opportunity as he would have in an unselected group. This argument would be more important if the homogeneous intellectual grouping were to extend to the playground, the gymnasium, the auditorium, and the social organizations. Since this grouping is only for the classroom, the objection need not be considered, except in so far as it affects the work of the class. Experience has demonstrated that in a homogeneous group, classified on the basis of a test, there are still many recognizable differences of ability, and that the rivalry for the leadership of one's peers is keener than for the leadership of a miscellaneous group.

Another objection is raised by those who feel that the slower pupils need the presence of the more rapid as a stimulus. Here, again, the lack of absolute uniformity furnishes in actual practice all of the stimulus necessary. In fact, it is usually more effective to have a pacemaker who is not too far in advance. Dozens of men were brought before the writer, while in charge of psychological examinations in a U. S. Army camp, accused of being stubborn and unwilling to try to perform their duties, while the real difficulty was that their pace makers were so far ahead of them as to be almost out of sight. When these men were placed in a group of their equals, with an instructor who understood their gait, real interest and competition arose among them, and the entire group moved forward at a much more rapid rate than they would have moved if left in a miscellaneous group.

The experiments so far conducted give little support to the objection that bright pupils when grouped together tend to overwork and break down. "Break down" from study is very rare, and when it does occur is more often due to trying to keep up with a group of more able pupils than to any other cause. "Overwork" is much more often "late hours" and "social life" than school work. It is not probable that pupils will really overwork when moving forward with other pupils of the same ability at their optimal rate.

The expectation that pupils classified in the slow moving group would feel the stigma of not being in the normal or rapid groups does not seem to be borne out by experience. It is true that where it is known that a given class is slow in its studies, and where the teachers have not been led to recognize that persons of "different" gifts from their own are nevertheless just as worthy, some few pupils have pointed a scornful finger at the "boobs," but usually without any serious consequences. The slow pupils are usually happier than under the miscellaneous grouping plan, and in many cases an unusual amount of class spirit has developed among them, possibly as a "protective reaction." It is certainly desirable, however, for the pupils and teachers to rid themselves of any feeling that the rapid group is deserving of any more honor and respect than the slow. The pupils should as far as possible know only

that they are in Miss B's or Miss E's room, without being informed of the real reasons for their assignments, except in special cases. Neither the pupils nor their parents have ever offered any objections to the homogeneous grouping as carried on at the Speyer School.

One of the greatest dangers now facing those interested in intelligence tests is that they will be accepted and used with too little critical judgment on the part of junior-high-school principals and other school administrators. It is so easy to become convinced that there is value in the method and so difficult to judge just how much dependence may be placed in it that many grievous mistakes are certain to be made. The same difficulty exactly arose in the U. S. Army cantonment in which the writer had charge of the psychological examination of troops. Company commanders, who were doubtful at the beginning, came to put entirely too much confidence in the results of the intelligence ratings of their new men.

An illustration of this uncritical attitude among well-trained school administrators was found by the writer in the Speyer Junior High School of Teachers College, in which homogeneous grouping has been most carefully practiced since 1915. Because of the greater inconvenience of scoring and tabulating the separate tests which had been used in previous years, the principal decided to employ the Otis Tests as the basis for his grouping of new pupils entering in September, 1920. Looking through the Manual for these tests, he found convenient "coefficients of brightness" which seemed to be worth more than the raw scores for his purpose. The pupils were therefore tested by the Otis Tests and their names arranged in order according to their coefficients of brightness. All pupils having "coefficients of brightness" from 241 down to 162 were placed in one section, those from 159 to 138 in another section, and so on for the five sections of the entering class.

The writer, having a group test of intelligence which he wanted to evaluate, asked permission to try it on the junior-high-school pupils and was surprised at the confidence with which teachers gave him information regarding the coefficients of brightness of their pupils. When the results of the new group test, the Mentimeters, failed to correspond with the Otis Coefficients, it was proposed to

the principal that still a third group test, the National Scale A, be given to these same pupils. When the results of the National Scale A failed to agree fully with either of the two previous tests, the principal began to ask which of the three tests came nearer the truth.

In order to determine the relative merit of the three tests in predicting the success of junior-high-school boys in this particular school, correlations were made (by the product-moment method) between the scholarship marks of these 120 pupils at the end of the first semester and their scores in each of the three intelligence tests. In the case of the Otis Tests, the correlation was higher with the coefficients of brightness than with the unmodified Otis scores, showing in our opinion, that the teachers' marks were influenced more decidedly by the derived ratings which they knew and upon which the pupils had been classified than by the relative abilities of the pupils. The coefficients obtained were as follows:

Scholarship marks and Otis C. B.'s	$r = .535, \pm .047$
Scholarship marks and Mentimeter Scores	$r = .481, \pm .050$
Scholarship marks and Otis Scores	$r = .470, \pm .050$
Scholarship marks and National Scale A Scores	$r = .459, \pm .051$

In order to determine the relationship of the three group tests of intelligence to each other, intercorrelations were made between the tests, with resulting coefficients as follows:

With	Otis C. B.	Otis Score	National Score
Otis Score	.851, ± .025		
National Score	.565, ± .043	.546, ± .044	
Mentimeter Score	.587, ± .040	.641, ± .037	.731, ± .031

The highest relationship between two tests was clearly between the National Scale A and the Mentimeter scores.

To determine the degree to which each of the three tests is a measure of language ability, the same pupils were given the Briggs Analogies Test Alpha. Its correlations with the scholarship marks and the three intelligence tests were as follows:

With Otis Test Score	$r = .442, \pm .050$
School Marks	$r = .419, \pm .047$
National A Scores	$r = .331, \pm .055$
Mentimeter Scores	$r = .297, \pm .059$

It would appear that the scores in the Otis Tests were influenced by language factors, and that the scholarship marks were influenced by the same factors. It would not be most economical of time, therefore, to give both the Otis Tests and the Briggs Analogies Test, for they are too nearly alike. Economy would suggest combining two tests which correlate little with each other, but highly with school success, thus getting as wide a range of different intellectual abilities as possible to use as a basis for homogeneous grouping.

Examination of the foregoing correlations and of the correlations of the individual tests contained in the three test booklets led to the conclusion that the Otis C. B.'s were less satisfactory as a basis of homogeneous classification for these particular boys than the Otis Scores would have been, and that the Otis Scores were less useful than the scores of either of the other two tests would have been. In the case of older pupils or of younger pupils, or in the case of junior-high-school pupils in other places, it is possible that the relative value of the three tests would be changed. It is also possible that the relative value of the tests would be different in this same school if the purpose were something other than the prediction of school success in the first year of junior-high-school work. Actual trial is the only safe method of determining the value of a test for a given purpose, and one should not be satisfied with a test which works fairly well if another can be found which works better.

One of the characteristics which experience has indicated as necessary in a satisfactory group test of intelligence is that the separate tests composing it should be steeply graded in difficulty from easy to hard, and that the time limits be so adjusted that one's score will indicate *how difficult* a problem can be solved, to a greater extent than it indicates *how many* he can solve in a given time. *Speed* tests are less indicative of ability to do school work than *power* tests. The dullest pupil must make a considerable score and the brightest pupil must not approach a perfect score if the test is to indicate relative strength with anything like precision. For the classification of junior-high-school pupils, therefore, the tests composing the battery should each be so easy at the

beginning that second or third-grade pupils could make some appreciable score and so difficult at the end that college students could not make perfect scores.

Summary

Intelligence tests have been used successfully in the educational guidance of pupils of junior-high-school age and in the classification of such pupils into groups of homogeneous intellectual ability. The evidence they furnish should be supplemented by all of the exact information it is possible to secure about each pupil, and these data should be evaluated by someone who uses good "common sense" and understands the limitations of the tests and of the other evidences. Changes of classification should be made promptly whenever new evidence is found that outweighs the data upon which previous action was based.

The classification of junior-high-school pupils into groups having common educational and vocational goals, and into subdivisions having the same ability to make progress toward these goals, is only the beginning of the real problem of adjusting the school to the abilities of its pupils. Homogeneous classification is not an end in itself. Teachers must be brought to recognize the usefulness and dignity of the classifications and must be trained to advance each group at its optimal rate. Administrators must be constantly on the alert to find the best means possible for the classification of their pupils and should not be tempted into the acceptance and use of a scheme without scientific evidences of its superior value.

CHAPTER VII

THE ADMINISTRATIVE USE OF INTELLIGENCE TESTS IN THE HIGH SCHOOL

W. S. Miller
Professor of Educational Psychology, University of Minnesota, Minneapolis, Minnesota

In 1914 the writer, under the direction of Dr. Whipple, began the preparation of a thesis on "Mental Tests and the Performance of High-School Students as Conditioned by Age, Sex, and Other Factors." It was hoped that as a result of the investigation a battery of tests might be developed that could be given to groups of high-school students, thus providing the principal or superintendent with a convenient instrument for predicting probable success in high-school work. At that time no such instrument had been developed. Furthermore, practically no reliable norms had been established for single tests that might be used in such a battery of tests.

In this thesis the value of a group test was emphasized, and in the closing paragraph it was predicted that in the near future (within a half-century) the mental testing of high-school pupils would be as common as physical examination is in the larger and more modern high schools.

The writer could not have foreseen psychological examination in the army, with its resulting impetus to mental testing in the public schools, as a result of which within a decade mental testing has experienced a growth and development which normally would have required a much longer period.

In general, this rapid growth has been advantageous and fortunate. It is true, however, that the testing movement is likely to suffer from 'growing pains' and to receive some reverses on account of this rapid development. Psychologists have been marketing group tests at a rapid rate, some of which under normal conditions

would have been tried out more thoroughly before placing them in the hands of school administrators only partially trained in administering them. More significant, however, is the fact that school administrators and teachers have not had the opportunity for securing training in the use and interpretation of the tests. As a result of this lack of information school administrators and teachers who have not studied the movement are dividing into two camps. Those who are by nature skeptical can see no value in attempting to measure anything so complex as general intelligence. They see in mental tests another educational fad and are willing to treat them as such. The other camp, a more credulous group, accepts mental tests as a mysterious instrument with which they are able within a period of thirty minutes to judge a high-school pupil's value to human society. They are believers, although too often they do not know clearly what they believe. Those who want to see the full value of mental testing realized sometimes can not help wishing that these believers were less credulous and enthusiastic.

School administrators and teachers who have made a careful study of mental testing see in it little that is really new except the scientific method by which it is done. They realize that for many years superintendents, principals, and teachers have questioned students and by their answers have formed judgments of their ability to succeed in school work. They see in mental tests an instrument for supplementing their crude and hasty judgments. They realize that mental tests are not infallible and that many conditions may modify a test score, making it misleading and unreliable. They know the degree of reliability of the tests and govern themselves accordingly. They realize how difficult it is to judge accurately the general intelligence of a high-school pupil and therefore welcome mental tests as an aid which furnishes within a short period of time objective data that make comparisons fairly reliable.

The author (as principal) has had an opportunity to observe these attitudes among the teachers in the University of Minnesota High School, where for the past five years pupils have been tested and classified on the basis of the results of the tests alone. A sane attitude toward tests develops as the knowledge of the possibilities and limitations of tests develops.

These same attitudes were manifested by officers in the United States Army. The mental tests were of greatest service among those officers who realized their possibilities and limitations. The officer who wished to get rid of a subordinate officer with fifteen years' experience because he rated "C" on the Army test did not understand that fifteen years' training of an average man in a relatively simple mechanical activity would give service quite comparable to that of a high-grade man trained in the same field for a period of two or three months. Officers failed frequently to comprehend that the tests did not give a measure of all the desirable virtues a man might possess. The tests were designed to measure general intelligence only and could not for that reason measure the results of specialized training. Every psychological examiner in the army was confronted first with the problem of educating those who were to make use of the tests in order to prevent their misuse. Similarly, the problem of the proper use and interpretation of tests of high-school pupils embodies a problem of education in view of the fact that the giving of the tests, the administrative use to be made of them, and their interpretation are in the hands of men and women with little training in the field of mental tests. It is encouraging to note in this connection the large increase in enrollment in courses in educational psychology and mental tests in our colleges and universities, especially during the summer session. Educational periodicals are rendering excellent service in this educational program. The officers of the National Society for the Study of Education are to be commended for devoting their entire *Yearbook* to the discussion of intelligence tests.

What Do Mental Tests Measure[1]

Mental tests are designed to measure native mental ability, not achievement. The school administrator should not confuse mental tests with achievement tests. They serve quite different functions.

[1] For a full, and somewhat technical discussion of this complex question read "Intelligence and its measurement: a symposium," by E. L. Thorndike, L. M. Terman, F. N. Freeman, S. S. Colvin, Rudolph Pintner, B. Ruml, S. L. Pressey, V. A. C. Henmon, Joseph Peterson, L. L. Thurstone, Herbert Woodrow, W. F. Dearborn, and M. E. Haggerty. *Journal of Educational Psychology,* 12: March and April, 1921.

The achievement tests are designed to measure the results of a pupil's attempt to master a definite field of knowledge. They attempt to tell how successful his efforts have been. The mental tests are designed to tell, in advance of any effort, how well the pupil would succeed if he attempted to master a definite field of knowledge. Achievement tests are a measure of what *has happened.* Mental tests measure native ability, which is one important factor in predicting what *will happen.*

One frequently hears it said that the results of mental tests are almost wholly dependent upon the previous training of the person tested; in other words, they are thought of as achievement tests, the results of which show, not native ability, but the presence or absence of favorable environmental influences. It is doubtless true that mental test results do reflect the influence of the environment of the pupil tested; but we may ask, to what extent is the mental test score determined by environmental factors? Are environmental factors so potent that they render the test score useless as an index of native ability, or are their influences so slight as to be almost entirely disregarded? A child reared in an environment where, despite his desires, he was not taught to read, would of course score zero on a test designed for literates. Obviously, his score would in no sense be a test of his native ability, but rather a test of his reading ability. This illustration makes it clear that in making mental tests it is necessary to assume a minimal *common* environment for those who are to take the test. In constructing a test for high-school and college students one is justified in assuming literacy of the average fifth-grade child. To reduce further the errors that might arise from variation in speed in reading and writing, the amount of reading and writing required in the test is reduced to a minimum. With these precautions in the selection of test material suitable to the group to be examined, it is not likely that differences in environment within the group would invalidate the mental test scores. The examiner should, however, take account of extremely unfavorable environmental factors in individual cases, for example, language deficiencies of foreign pupils, and re-examine them with tests that do not presuppose ability to read English.

Mental tests containing general information, arithmetic problems, opposites, and vocabulary are condemned by the layman as tests of mental ability because they are unfair to pupils with unfavorable school, home, and social influences. If pupils exposed to unfavorable environment always did poorly in these tests, the objection might be more significant. Even then one would have to reckon with physical inheritance as well as with social inheritance.

Furthermore, children of approximately the same age, reared in the same home, taught by the same teachers, may receive radically different scores on these tests while children from most contrasted environments may receive similar scores.

Some raise the question of the time limits on the tests, which they say, make the tests unfair to the "slow, accurate thinker." Experimentation[2] has shown that doubling the time on the Army Alpha makes very little difference in the relative standings; the coefficient of correlation between scores based on standard time and scores based on double time is 0.965. The median score of the group that had double time was, of course, higher; but the relative position of the men was practically unaltered.

Contrary to general belief, the slow thinker is not necessarily the accurate thinker. This can be demonstrated by selecting one group of test papers in which only 50 percent of the items are attempted, and comparing the accuracy of this group with another group of test papers in which 75 percent of the items are attempted in the same period of time. Although the opportunity for error in the latter group is 50 percent greater than in the former, it will be found that the rapid pupils have a smaller percentage of error than the slower pupils.

Some school administrators contend that physical and mental conditions fluctuate so much from day to day that mental tests can not be relied upon as a measure of a pupil's general intelligence. It is true that extreme physical or mental disturbance at the time of an examination may materially alter the mental test score of an individual pupil. If these abnormal conditions are known, the examination of the student should be postponed. The unreliability of tests due to abnormal physical and mental conditions may be

[2] *National Academy of Science Memoirs*, 15: 1921, Part II, Ch. 9, p. 416.

almost entirely eliminated by repeating the same test with a week intervening, or by giving different forms of the same test or by giving different tests and using the average of the two trials.

The question what mental tests really measure is of general interest to the school administrator but the question he is more interested in from a practical point of view is; do mental tests enable the administrator to predict success of a pupil in high-school work? This question will be answered in the section, "Mental Tests and School Marks."

The Selection and Giving of Mental Tests

School administrators will experience little difficulty in selecting high-school tests, since the psychologists in making the tests usually have the administrative use of the tests in mind in their construction.

A good test for high-school students should meet the following standards:

1. The test should differentiate. It should be sufficiently difficult to test the most capable pupil and easy enough to permit the least capable pupil to do something with it. In brief, the results of the test should contain neither zero nor perfect scores.

2. It should possess a high coefficient of reliability. The coefficient of correlation between two applications of the test should be above + 0.80. The higher the coefficient of reliability, the better.

3. It should give a coefficient of correlation of + .50 or higher with average school marks and with the estimate of intelligence of pupils by teachers. In applying this criterion it should be kept in mind that unreliable marks and poor judgment of teachers may be factors in lowering the correlation.

4. The instructions for giving the test should be simple and direct. The technique of giving the test should not be complex.

5. The directions to the pupil should be such as to insure a clear understanding of what is to be done in the test. Ample fore-exercises aid in obtaining a clear understanding by the pupil.

6. The test should be so constructed as to make possible, rapid objective scoring.

7. It is convenient to have the time needed for giving the test limited to a single high-school period of forty minutes.

8. It is not necessary to call attention of administrators to the fact that cost is one criterion that should not be overlooked.

All tests for high-school pupils now available are accompanied by a carefully prepared manual of instructions for giving the tests. It is imperative that administrators follow these instructions *verbatim* and that the giving of the tests be entrusted only to such persons as understand the importance of uniformity in method of giving tests. Comparison of groups within the school system and comparison with standard norms will mean nothing unless uniformity of method of giving the test is secured.

Where assembly halls are available, a large number of pupils may be handled by a single examiner with an adequate number of proctors.

Seats with arms on which to write are desirable; but where these are lacking, lap boards are a convenient substitute. In so far as possible, pupils should be so seated as to remove the temptation to copy.

Proctors should make notations on the papers of individual pupils who suffer interruptions or exhibit irregularities that would clearly modify the test score, such as copying, illness, improper attitude, confusion in turning to next test, and lack of effort.

The work of scoring mental tests is not particularly irksome when it is done promptly and systematically by all of the teaching staff. Speed and accuracy are secured by assigning one teacher or a group of teachers to a single test. They soon learn the key and the whole process becomes relatively automatic. The addition of the separate test scores should be assigned to a teacher who is rapid and accurate in the process of addition, and the additions should be checked by another person if an adding machine is not available. Another teacher should be assigned to classifying scored tests according to sex, age, grade, etc.

By a systematized procedure the staff of a high school of 400 pupils could score any group test for the entire school in from two to five hours. By a haphazard procedure the same task might worry an entire staff at odd intervals for a week or more. Admin-

istrators reading this will in many cases be reminded of piles of unscored tests in their offices that have not received this prompt and systematic treatment. Let us see to it that tests are not placed on the shelf along with unused laboratory equipment, purchased because it was fashionable and well advertised. Tests are of no use until they are scored, but much remains to be done after they are scored.

Recording the Test Scores

The author examined all entering pupils in the University High School for four years before providing for a satisfactory record of the results. If the test scores are to be of value they must be readily accessible to teachers and administrators. The place for the test scores of individual pupils is on the permanent record card, which should contain among other things the pupil's scholarship record for the four years. The following is suggested as a convenient form for the mental test record on the permanent record card.

Name of Test	Date Given	In what Grade	Standard Median	Score	Class Median	Percentile Rank in			I.Q.	E.Q.	I.B.
						Standard Scores	Class	School Marks			

The date should be included because the interpretation of a test score obtained in the freshman year would not be the same as that of one obtained in the senior year. The percentile rank (P. R.) gives the score a meaning in relation to a large group. Percentile rank may be interpreted as the percent lower. This will be discussed later on. Intelligence quotient (I.Q.) provides a rating which makes allowance for the age of the pupil. Some group tests provide approximate I. Q. ratings. Where data are available, the efficiency quotient (E.Q.) could be recorded.

The reasons for placing the mental test record on the permanent record card are so obvious that they do not warrant extended dis-

cussion. Interviews with pupils in regard to scholarship may be made more intelligently with knowledge of their standing in the mental tests. Both records are available at once by this method. By having the test records on cards the calculation of coefficients of correlation is simplified.

Only once in the author's experience has he received a record of mental tests on a transfer credit blank. In this case "P. R. 38; I. B. 91" was written at the bottom of the card. This suggested that it would be advisable to provide adequately for a mental test record on the blank for transferring credits. This is important since it gives an official record of the tests the pupil has taken, thus making duplication of tests unnecessary. If the pupil is given the same test twice, the second score may then be interpreted in the light of his previous experience with the test. The form of record on the transfer credit blank could very well be a duplicate of that on the permanent record blank.

Tabulation of Results

Age-Grade-Score Distribution

For convenience in the tabulation of the results of testing 6000 high-school pupils in Minnesota the author devised a blank[3] which shows the distribution of scores for all ages for grades 7 to 12. The instructions for the use of the blank are printed on the back of the blank. This is a convenient device for collecting data for graphs like those in Figs. 2 to 9. It serves a triple function as a tabulation sheet, a percentile graph, and a correlation graph (See Fig. 1).

The figures in the vertical column at the left (Fig. 1) represent the units of the Miller test score by tens. The figures at the head of the other columns are the intervening 9 digits. The figures at the bottom will be explained later.

Let us assume we wish to tabulate the results of the tests of a ninth-year class of 80 pupils. We will use the dot (.) as a tally

[3] Published by the World Book Company, Yonkers, N. Y.

symbol. It will be convenient to have one person read the scores and another do the tallying, although one person can do both. Assuming the first score read to be 83, a dot would be placed in the column headed "3" to the right of "80" in the left-hand column.[4] A score of 37 would be indicated by a dot in the column headed "7" to the right of "30." A score of 20 by a dot placed in the column

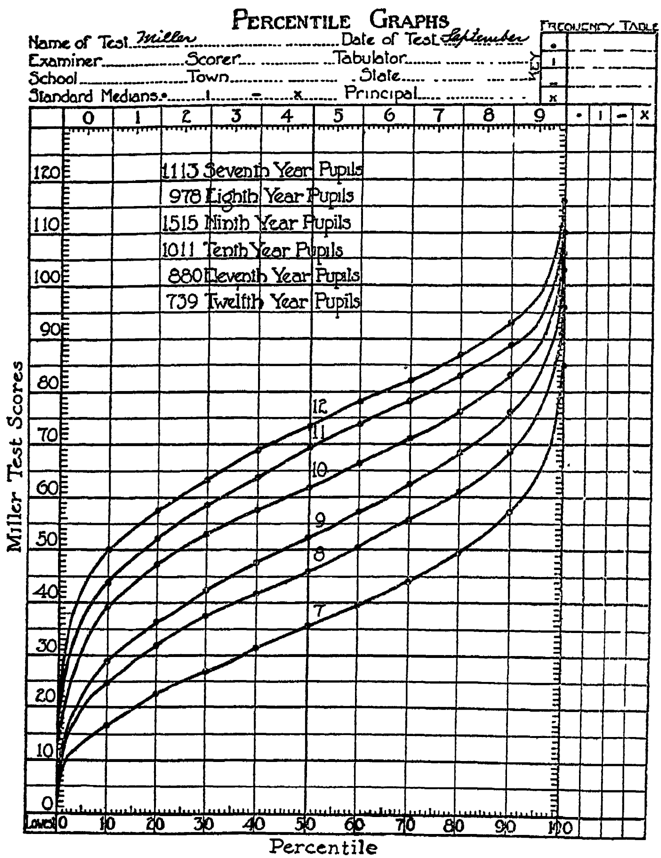


FIG. 1
By courtesy of the World Book Company, Yonkers, N. Y.

[4] The use of the blank as a tally sheet is not illustrated in Fig. 1.

headed "0" to the right of "20." It will be observed that this method locates each score to the smallest unit of the scale.

When all the 80 freshmen scores have been tallied, a table of frequency by tens may be made by counting the dots horizontally across the blank for each ten units and placing the number at the proper level in the column immediately to the right of the column headed "9," the column headed by a dot.

Three other classes may be tallied in the same way on this same blank by using the other symbols indicated in the key. Write after each symbol in the key the name of the group it represents.

The Percentile Graph

As an aid in tabulation and to facilitate the interpretation of the results of tests the percentile graph will be found most convenient.

In constructing a percentile graph of the 80 freshmen scores, locate the lowest score made by a freshman. Let us assume that the lowest score made is 23. Make a small circle, (o), on the scale at the left, on the vertical line rising from the zero percentile, at 23. The next point on the graph will be the score of the freshman who is 10 percent of the group above the lowest. Since there are 80 in the group, the tenth percentile would be the eighth freshman. Beginning with the lowest, count the tallies in order to the eighth. Note what the score of the eighth freshman from the lowest is and put a small circle at that point on the vertical line locating the 10th percentile (marked 10 at the bottom). The twentieth percentile score would be that of the sixteenth freshman from the lowest; the thirtieth percentile, the score of the 24th freshman, etc.

When the remaining percentile scores have all been indicated as was explained for the tenth and twentieth, join the small circles by a curved line.

Percentile graphs for the other three classes may be constructed in the same manner on the same blank. There are shown in Fig. 1 percentile curves for students of six different school years.

If one does not wish to use the blank as a tally sheet, data for the percentile graph may be obtained by stacking the test papers in order from the lowest to the highest. Then the several percentiles

may be located by counting through the papers, noting the score found on the test paper that represents every tenth percentile.

The graph shows the range of scores from the lowest (lower left) to the highest (upper right).

The point where the percentile graph crosses the 50th percentile line locates approximately the median for the group and may be read directly from the scale on the left. (See Fig. 1.) The 25th and 75th percentiles (the first and third quartiles) of the group may also be located in the same manner as the median by reference to the graph.

To determine the percentile rank of any individual freshman proceed as follows: Locate his score on the scale at the left; from this point follow an imaginary horizontal line to the point where it intersects the percentile graph for the ninth year; from this point of intersection let fall an imaginary perpendicular. The point of intersection of this perpendicular and the base line is his percentile rank, P. R. This figure shows the percent of the group that is lower than this individual.

One common method of comparing two groups of pupils is to state the percent of one group that falls above or below the median of the other group. For example, in Fig. 1 find the median of the freshman group (intersection of 9th-year curve with 50th percentile); follow an imaginary horizontal line to the left to the point of intersection with the percentile curve for seniors. From this point let fall an imaginary perpendicular. The point of intersection with the base line will be the percent of the senior class that is below the median of the freshmen class. The percent of seniors above the median of the freshmen is 100 minus this number.

The results that appear in the percentile graphs which follow make it evident that the score of a pupil of any given age should be interpreted in the light of the grade location of the pupil. For example, from the percentile graphs for pupils 16 years of age, Fig. 5, it will be noted that a pupil 16 years of age in the seventh year, scoring 55 would have a percentile rank of 95, in the eighth year a percentile rank of 88, in the ninth year, 66, in the tenth year, 26, in the eleventh year, 17, and in the twelfth year, 0, *i. e.,* 55 is the lowest score obtained by any pupil 16 years of age in the senior year in high school.

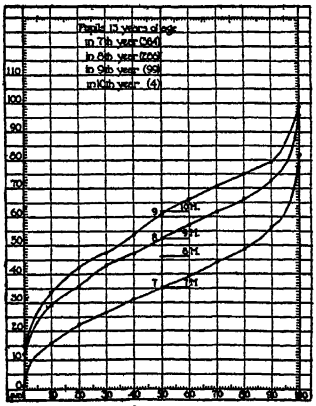


Fig. 2

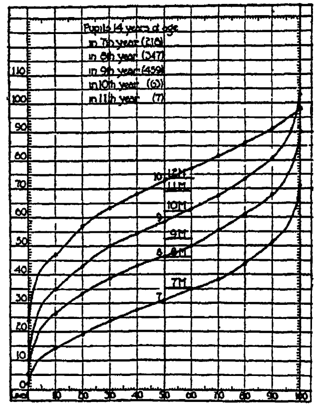


Fig. 3

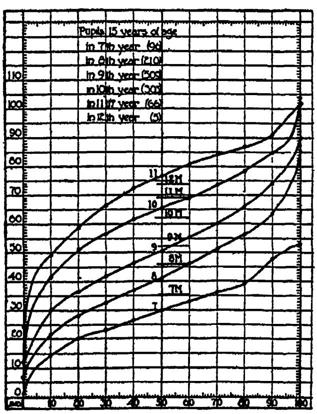


Fig. 4

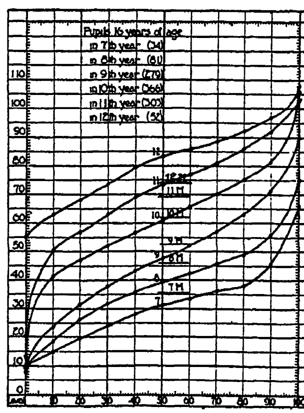


Fig. 5

The same score, 55, interpreted in the light of norms for pupils of all ages in grades seven to twelve (Fig. 1) would show the pupil to have the following percentile rank; in seventh year, 88; in eighth year, 68; in ninth year, 56; in tenth year, 33; in eleventh year, 24; in twelfth year, 17.

With the explanation of percentile graphs already given, the reader should be able to interpret the percentile graphs without further detailed explanation. On each percentile graph the medians

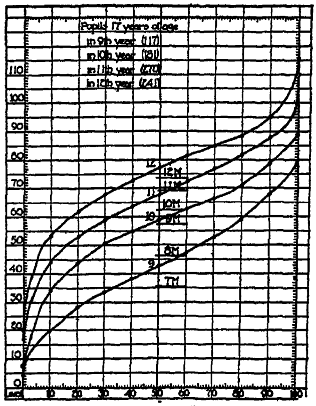


Fig. 6

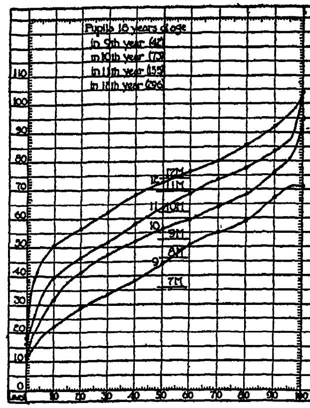


Fig. 7

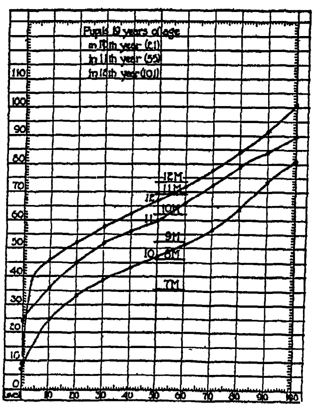


Fig. 8

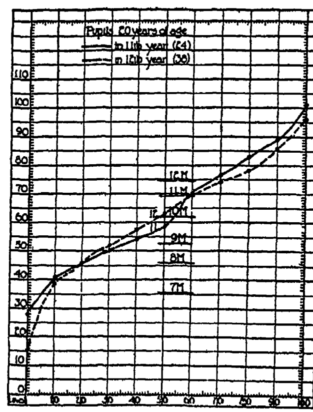


Fig. 9

for grades 7 to 12 are indicated by short lines on the 50th percentile. It will be observed in Fig. 5 that the medians for pupils 16 years of age in the seventh, eighth, and ninth grades are below the standard medians for those grades. The median for pupils 16 years of age in the tenth grade is almost the same as the standard median for that grade. The medians for pupils 16 years of age in the eleventh and twelfth years are above the standard medians for those years.

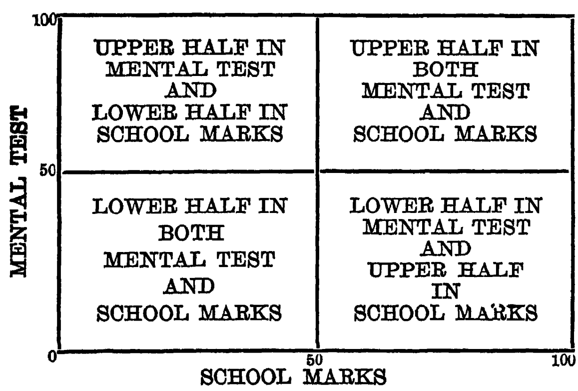
Correlation Graphs

The percentile graph blank (See Fig. 11) is very convenient for showing graphically the correlation between test scores and school marks, or the correlation between the different mental tests.

To construct a correlation graph on the percentile graph blank first convert the test scores and school marks into percentile ranks. The percentile ranks may be obtained with a fair degree of accuracy directly from the percentile blanks as already explained.

In the correlation graph indicate the position of each pupil by a small circle. A pupil with a percentile rank of 90 in the test and a percentile rank of 80 in school marks would be located at the intersection of the horizontal line marked "90" with the vertical line marked "80", assuming that the percentile ranks in the test are plotted on the ordinates (the verticals) and the percentile ranks in school marks are plotted on the absissae (the horizontals).

The fiftieth percentile lines in the tests and school marks divide the graph into quarters. It will be observed that all pupils in the different quarters may be described as follows:



Classification on the Basis of Test Scores

The percentile graphs of Fig. 1 show the wide range in scores in any one year and also the overlapping of all of the years from the seventh to the twelfth. The fact that high-school students vary widely in ability was known long before any one thought of using

mental tests. It is true, however, that in spite of this knowledge we have continued to try to teach all pupils the same material by similar methods in the same period of time. Experience has shown most administrators many times that high-school pupils can not be handled satisfactorily when treated as if they were a homogeneous group. This has led to numerous administrative schemes intended to take care of these individual differences. The tendency among administrators is and has been to put too much faith in the device without enough attention to the actual teaching process.

In schools that are large enough to have more than one section in any given subject, much can be gained by sectioning the pupils on the basis of the mental test scores.

For five years the entering freshmen in the University of Minnesota High School have been given mental tests prior to the opening of school. The class is large enough to make only two sections. Those above the median in the tests are assigned to one section and those below the median to another section. At the time they are given the mental tests they are asked to fill out class cards for each subject they wish to take, leaving blank the room, period, and section, which are filled in by the office secretary after the tests have been scored. The pupils are asked to call at the office for the cards on the opening day of school. These class cards provide the pupils with their schedule of classes and serve as admission cards to classes. The teacher collects the cards and has at once her class roll. The same plan of registration is followed for the upper classes, except for the mental tests, which were given when they were freshmen. They fill out the class cards at the close of the preceding year. This plan of registration gives the principal control of the segregation of pupils of like destination or like program, thus avoiding overcrowding of certain sections, conflicts, and the general confusion that is so prevalent during the opening days of a high school. This is not the place for a detailed discussion of program making. High-school principals should read Mr. Richardson's monograph[5] dealing with that problem.

---

[5] Myron W. Richardson, *Making a High-School Program.* School Efficiency Monographs, World Book Company, 1921.

Experience with division of a class into two sections reveals the fact that even greater advantages would be derived from a division into more sections, as would be possible in a larger high school. With a larger number of sections each of them would be more homogeneous in ability. A freshmen class divided into two sections still shows a wide range of ability in each section—too wide, in fact, for the most effective work.

Classification of high-school pupils on the basis of mental ability results, or should result, in certain advantages:

1. *It makes possible an adaptation of the technique of instruction to the needs of the group.* It makes possible such an adaptation, but it does not insure it. The tendency too often is to use exactly the same method for the different sections. Unfortunately, we do not yet know enough about differences between methods of instruction for, let us say, the upper tenth and the lower tenth. It is generally recognized that less capable pupils require much more detailed explanation than the more capable, and that the former require much more drill to make certain skills automatic than do the latter. It is not to be expected that the teacher's preparation or presentation would be the same for all sections. Classification alone will not bring the results desired; it is only a means to an end.

What progress of a class as a whole may we expect when each individual in a heterogeneous group is given the same task with the same period for its accomplishment? Measured results show that the ratio of the poorest to the best student in a class is often 1 to 8 when the task assigned is reproducing ideas gained from reading a paragraph. If, for example, a lesson of this sort were assigned with one hour for preparation for the best pupils, it would be reasonable to expect that it would require 8 hours for the poorest pupil to prepare the same lesson equally well. If, on the other hand, a lesson were assigned which the poorest could prepare in one hour, the best pupil could prepare the same lesson in less than 8 minutes.

With this wide range of ability it might be suggested that a lesson of such length should be assigned that the median pupil could prepare it in one hour. Preparation of this lesson suited to the median pupil would require four hours by the poorest pupil;

while the best pupil would prepare the same lesson equally well in less than half an hour.

To illustrate further the difficulties of group instruction with pupils that vary widely in ability, let us imagine the poorest pupil in the analogies test sitting in an algebra class beside the best pupil in the same test. The analogies is a test of speed in perceiving logical relations; it shows a significant positive correlation with performance in algebra, also with the teacher's estimate of general intelligence. In a class to which the author gave the analogies test as an individual test, the best pupil could perceive the relation and speak the missing word at the rate of one in each 3.5 seconds; the poorest pupil could perceive the same relations at the rate of one in each 27.4 seconds.[6] Let us designate the best pupil "B" and the poorest pupil "P." Let us suppose that in order to progress understandingly with the work in the recitation it would be necessary to perceive relations at the rate of one every 10 seconds. "B" would perceive relation No. 1 in 3.5 seconds and wait 6.5 seconds for relation No. 2, but "P," if he were not distracted by the appearance of relation No. 2 would require 27.4 seconds to perceive relation No. 1. By the time "P" has grasped relation No. 1, it is almost time for relation No. 4, but the perceiving of relation No. 4, let us assume, is dependent upon his having grasped relations No. 2 and 3. It is evident that the recitation would not continue long at this rate before "P" would be hopelessly lost; while "B" would be bored by the tedium of waiting for each succeeding relation almost twice as long as it took him to perceive the relation when it was presented. With the knowledge of the abilities of "B" and "P" which the analogies test affords, it would not take a wise man to predict that, if "P" were held to a standard adapted to "B," he would fail to gain credit in the course. If, on the other hand, the recitation progressed at a rate suited to "P," "B" would lose interest and the recitation would fall far short of calling forth the best that was in him. Who can estimate the deadening influence on "B" of four years of high-school work on

[6] In giving the test, the pupil was allowed no more than 30 seconds for each analogy. If the correct answer was not given in 30 seconds, the time was recorded as 30 seconds. This average is therefore less than the actual time required to see the relation.

this level? What can we expect from "P," who must of necessity be completely "muddled" at the end of each recitation?

If "P" is to make normal progress, he must be given more time to see relations and to answer thought-questions. This is not advisable if "B" is to participate in the same recitation. It would, therefore, seem advisable to place "P" in a class of pupils who would profit by the long interval that must elapse between question and answer, and to place "B" in a class of pupils like himself mentally.

The writer is convinced that in classes as organized at present thought-questions are put at a rate too rapid for a large majority of the class. The rate in most classes is more nearly adapted to the best 10 pupils in 100. Anyone may be convinced of the truth of this statement by observing teachers of freshmen classes in the high school if he will take the trouble to measure with a stop-watch the interval of time allowed for answers to thought-questions. The median time required by freshmen to see the simple relations in the analogies test we employed was about 14 seconds. Most teachers, especially beginners, show considerable uneasiness, at least, if answers to thought-questions that involve the grasping of relations much more complex than those in the analogies test are not forthcoming within 10 seconds. If the answer is not given almost immediately, the teacher interrupts by meaningless remarks, by a needless repetition of the question, by passing the question on to some other pupil, or by answering the question herself. She can't endure the silence that must prevail while the pupil is thinking and organizing his material, and commonly feels that she must break the silence by making a remark of some kind, however useless and distracting it may be.

During the past year the author has had occasion to observe the work of over 100 practice teachers. There was no one fault more common than the one under discussion. It is due to the failure to recognize the fact that time is required to perceive thought-relations and that a large proportion of the time in the recitation must be allowed for the exercise of this important function. Fourteen seconds seems a long time to wait for a student to see relations as simple as those in the analogies test, in which the relation when

perc[illegible]expressed by a single word and in the presence of one p[illegible]any of the thought-questions put by teachers are much more complex than that and necessitate framing the answer in good connected English and giving it before thirty of his classmates. If the reader is a teacher, he can observe this fault by putting a thought-question to some member of his class and then measuring with a stop-watch the interval that elapses between the question and the expected answers. It is rare, indeed, that the teacher does not show considerable uneasiness before ten seconds have elapsed.

Miss Stevens[7] has attacked this problem from a different angle—the number of questions put during a recitation. In the light of the foregoing discussion it is clear why there are reasons for alarm when it is reported that recitations are frequent in which 200 or more questions are asked.

2. *Classification makes possible but does not insure an adaptation of materials of instruction to the needs of the group.* It is probably only a question of time until the makers of textbooks will recognize the wide range of ability among students and will make texts adapted to the different groups. It is possible now to select texts in general science of varying degrees of complexity. Some of these texts are well adapted to students in the lower third in abil[illegible]are for most part a bore to the upper third who know [illegible] material contained in the texts before they enter [illegible]hool. The scientific interests of the superior pupils [illegible] be deadened by spending thirty-six weeks largely in [illegible] the words of that particular author.

[illegible]me criticism might be made of materials in English, agriculture, domestic science, American history, and beginning mathematics. Simplification of texts for students of mediocre or less ability is desirable and necessary, but not for those of superior ability. This should not be interpreted as a plea for textbooks that are obscure and complex, but rather a plea for materials that for most part are new to the superior pupil and sufficiently involved to challenge his ability.

[illegible] Stevens, *The Question as a Measure of Efficiency in Instruction.* [illegible]llege, Columbia University, Contributions to Education, No. 48.

A more comprehensive treatment of materials rather [illegible]ore rapid progress through the high school seems to me to be [illegible] solution of the problem of the superior pupil. If this is to be the solution, a more intelligent selection of materials is imperative.

3. *Classification may make competition operative as an incentive.* The capable pupil may be freed from the boredom that ensues from the snail-like progress that is necessary if the slower student is to profit by the instruction. Competition may become for him an incentive to real work. The less capable student, when segregated, experiences the thrill that comes from being first. "Better be first in a little Iberian village than second in Rome." In a fat man's race the participants manifest considerable enthusiasm and interest, which is likely to be lacking if an expert track man is entered. Competition between the fat man and the track man does not operate as an incentive. It is evident that the fat man suffers humiliation and embarrassment and that the track man, if he is a good sportsman, misses the thrill that comes from the defeat of a worthy adversary.

It is not uncommon to hear teachers, principals, and superintendents who have had no experience in working with pupils classified on the basis of ability, object to such classification on the ground that the students in the lower sections would become d[illegible] and would make no effort when deprived of the stim[illegible] superior pupil, but I have never heard this objection [illegible] teachers and administrators who have actually classified [illegible] the basis of ability. Instead of being discouraged, the le[illegible] pupils are encouraged to compete when they realize th[illegible] chance for them to do as well as their neighbors. It is true that the recitation does not move so rapidly, since it is impossible, when the recitation lags, to 'pass on' the questions to the superior pupil, as is so often done when the superior pupil is present. Such a procedure does keep something happening but it does not contribute much to the understanding or progress of the inferior pupil. The inferior pupils in a mixed class soon learn that the better pupils carry the load of the recitation and to avoid embarrassm[illegible] inferior pupils are satisfied to let them do it.

It should be emphasized once more that the classification of pupils on the basis of mental ability does not solve all of the problems incident to group instruction. Sections of pupils that are the same in mental ability will contain pupils that vary in chronological age, physiological age, previous training, temperament, conduct, special abilities, social and economic status, and moral standards. The members of any class, whether it is or is not made up of students of equal mental ability, will vary in these characteristics, but members of a class of equal mental ability will vary less in them than will those of a class of markedly unequal mental ability. For example, the section of pupils of superior mental ability would be more homogeneous as to chronological age than an unselected class, since the former would contain a majority of younger pupils. The latter would contain most of the over-age pupils. These classes would therefore be also more homogeneous as to physiological age than would an unselected class. The section of superior pupils would contain more pupils with good previous training, better dispositions, better standards of conduct, better opportunities socially and economically, than would the class of inferior pupils.

While classification on the basis of mental ability does not insure uniformity in all of these characteristics, it is evident that the variation would be very much reduced.

In some localities administrators will encounter objections on the part of parents to mental testing and to classification on the basis of the testing, just as they encountered objections to physical examination a few years ago. These objections must be met tactfully by educating the public to the advantages to be derived from a testing program. Nothing is to be gained in the beginning by emphasizing in the minds of the children the significance of the classification. The wise thing to do is to assign them without comment to the section to which they belong. Teachers especially should avoid comparisons of progress, industry, etc., before the pupils.

Mental Tests and School Marks

In discussing the correlation between mental tests and school marks it is necessary to consider the reliability of both tests and

school marks. One could not claim that the tests are an exact and reliable measure of general intelligence even if psychologists could agree on what general intelligence[8] is. The tests are probably a more reliable indication of what a pupil's achievement in school *should* be than are his marks an indication of what his achievement has been. Higher correlations between mental tests and school marks than are now obtained can not be expected until marks are based more exclusively on *achievement*. Terman has pointed out the danger of grossly perverting the test as a measure of general intelligence by modifying the test to increase its accuracy as a prediction of school marks. To quote Terman:[9] "If we wished to devise a test which would give the most accurate possible prediction of the class marks a given group of college students would receive, we ought to include in it measures of personal beauty, voice quality, bashfulness, willingness to cultivate the good graces of the instructor, etc."

Teachers and administrative officers can increase the value of mental tests as an instrument for diagnosis by making school marks a more accurate measure of actual achievement. It is quite natural for a teacher to let the mark indicate in part a pupil's industry, cooperation, courtesy, persistence, honesty, reliability, punctuality, and disposition; but when achievement and all of these other items are indicated by a single mark, it is very difficult indeed to ascertain to what degree it is a measure of achievement. This concrete case will illustrate: A parent who was accustomed to permit his son, a seventh-grade pupil, to assist him in some simple arithmetical calculations observed that he was slow and inaccurate in his calculations; he observed also that his marks in arithmetic were all above 90. The father, anxious to check up his son's school marks in arithmetic, applied the Courtis standard tests in arithmetic and learned that his son's achievement was very poor. In addition, for example, he was about a grade and one half below the standard for his grade. In consultation with his son's teacher concerning the inconsistency of the mark in arithmetic the teacher admitted that

[8] "Intelligence and its measurement: a symposium." *Jour. of Educ. Psychology.* 12: March and April, 1921.

[9] *Ibid.*

the son was rather poor in arithmetic, but pointed out that he was a *good* boy, courteous, cooperative, and reliable. The father thought no less of his son because he possessed these desirable virtues, but he did think less of his son's marks as a measure of his achievement in arithmetic.

No one would deny that these items which the teacher mentioned and others are important and that much would be gained by constructing a report card that provided for a rating of the pupil on these items separately, reserving the mark that is written after each school subject for the measure of achievement in that subject. A pupil may be courteous, honest, reliable, industrious, attentive, and persistent and yet make a very poor mark in algebra. Both mental tests and school marks will be more meaningful with such a differentiated rating. The parent would then know that the achievement in algebra was low and that it was not due to a lack of industry, cooperation, etc.

The testing movement and the system of reporting by the public schools would be benefited greatly by the formulation of some standard uniform marking system. When such a system is formulated and certain symbols defined and applied to achievement and other items separately, we may expect a higher correlation between mental tests and school marks, and have in addition a language of marks that teachers, principals, superintendents, and parents can use and understand.

The standard achievement tests involving reasoning furnish a more objective criterion for checking the mental tests as an instrument for prediction. They furnish an illustration of a rating of achievement alone. A pupil's standing on a standard achievement test is not influenced by the numerous personal traits that color the teacher's mark.

The diagram reproduced as Fig. 10 shows clearly that even when emphasis is placed upon marking on achievement alone, as is done in the University High School, it is not always the pupil of low mental ability that fails; it will be noted, however, in comparing the marks of the lowest quartile group with the highest quartile group, that the former has about eleven times as many F's as the latter. About one fourth of the pupils in the highest quar-

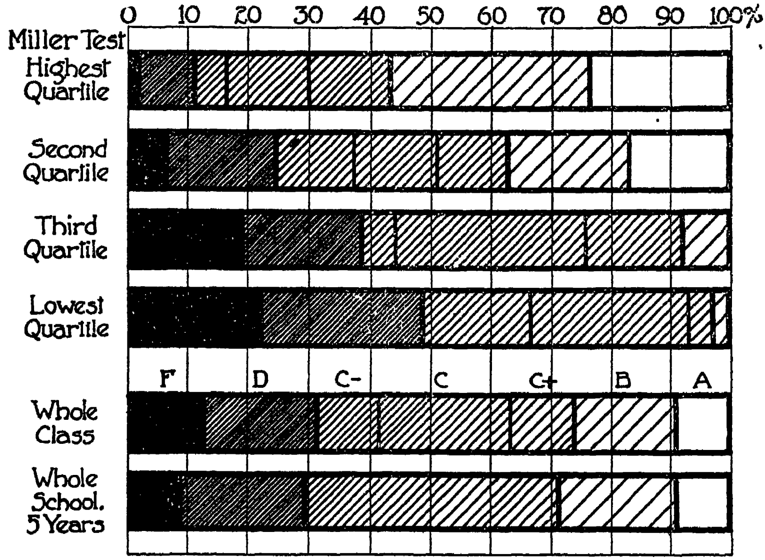


FIG. 10

Diagram showing the relation between the standings in the Miller Mental Ability Test and the average school marks (excluding gymnasium marks) of 55 freshmen University of the Minnesota High School, 1920-21.

tile received "A," while none in the lowest quartile received "A." The diagram shows clearly that mental ability as measured by the Miller Mental Ability Test is an important factor in determining the marks of high-school freshmen. The coefficient of correlation (Pearson) is + .522.

Administrators will find a graphic representation that shows each pupil's school standing in relation to his mental ability more useful for diagnostic purposes. The correlation graph, Fig. 11, furnishes this information in a form that is easily interpreted. Both the test scores and the school marks were converted into percentile ranks by the method already explained. The marks were weighted as follows: A, 100; B, 93; C+, 81; C, 69; C—, 50; D, 31; F, 7.

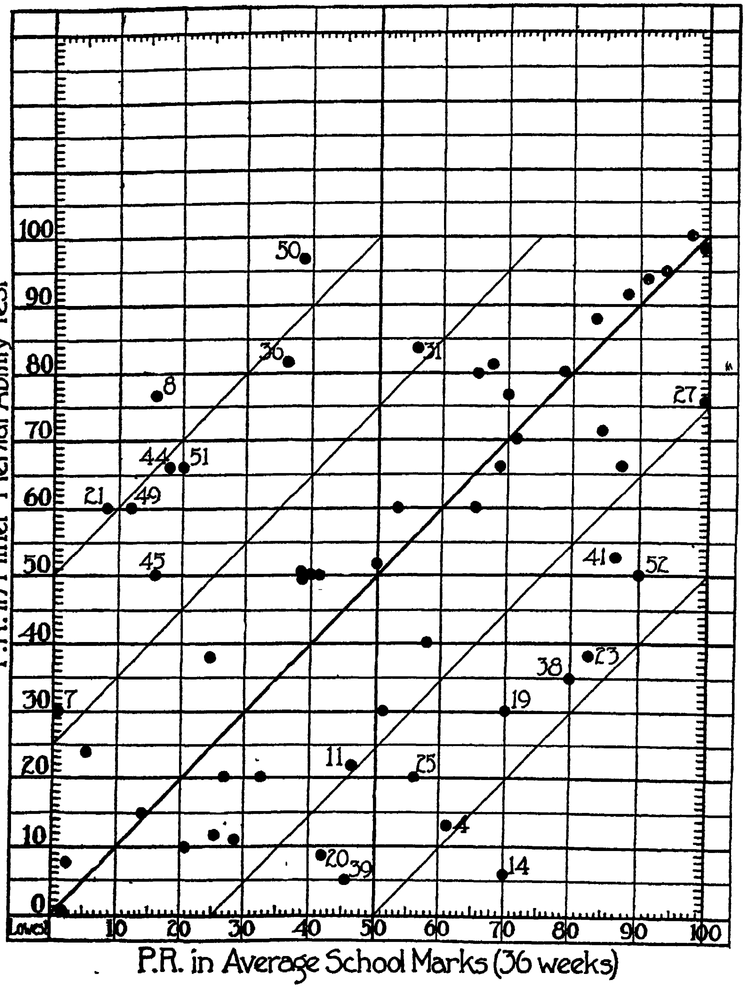


Fig. 11

If each pupil held the same percentile rank in school marks as in the mental test, the dots in the correlation graph, Fig. 11, would be on the heavy diagonal. Pupils whose percentile ranks in school marks and in the mental test differ by less than 25 points are be-

tween the diagonals originating at 25 on the horizontal and on the vertical scale. Pupils whose percentile ranks in mental test and school marks differ from 25 to 49 points are found between the diagonals originating at 25 and 50. Pupils beyond the diagonals originating at 50 differ in their percentile ranks in school marks and mental tests by more than 50 points.

Pupils at the right of the heavy diagonal hold a higher percentile rank in school marks than in the mental test.

Pupils at the left of the heavy diagonal hold a higher percentile rank in the mental test than in school marks.

Let us observe the facts concerning the relation between the test results and the school marks revealed in Fig. 11. It is obvious that the widest possible difference in percentile ranks in the two series would be 100 points, as would be the case with a pupil whose percentile rank in the test was 100 and whose percentile rank in school marks was 0. The widest difference found is 64 points (pupil number 14 on the graph). Four pupils, numbers 8, 21, 50 and 14, show a difference between percentile rank in the test and school marks of more than 50 points. Seventeen pupils, numbers 51, 44, 36, 31, 45, 49, 27, 41, 52, 23, 38, 18, 25, 39, 4, 7, and 20, differ in percentile ranks in test and school marks between 25 and 50 points. The remaining 34 pupils differ by less than 25 points in the two percentile ranks. The Pearson coefficient is + .522.

There are several factors that keep this correlation from being higher:

1. A test that can be given in 30 minutes and that involves only 19 minutes spent in actual work is not infallible as a measure of mental ability.

2. School marks are not, as every one knows, a measure of all a pupil is capable of doing.

3. School marks do not measure achievement alone. They are colored by courtesy, cooperation, industry, methods of work, previous training, etc., which the test does not measure.

It is interesting to study specific cases to ascertain the reason for the wider differences between percentile rank in the test and percentile rank in school marks. What are the chances that addi-

tional tests would show that this single test was unreliable as a measure of a pupil's ability?

In the University High School where the author was principal, the entering class (1920) of 55 members were given the Miller Mental Ability Test; Haggerty's Delta 2; Terman's Group Test of Mental Ability, Form A; Army Alpha, Form 8; Trabue's Mentimeters, and the Otis Test, in the order named. The first three tests were given on the same day, September 27, except for one half of the group who took the Miller test in July. The Army Alpha and the Trabue Mentimeters were given in October about two weeks apart. The Otis Test and the Stanford Revision of the Binet-Simon Tests were given in March, 1921.

The correlation (Pearson) between the Miller Test and the average of the first five tests given is + .903.

TABLE I.—55 NINTH-GRADE PUPILS, UNIVERSITY OF MINNESOTA HIGH SCHOOL
(All correlations in the table are positive)

	Delta 2	Terman Form A	Alpha Form 8	Mentimeter	Av. 1st 3 Tests	Av. 5 Tests	Grammar*	School Marks	Otis
Miller	.784	.747	.76	.768	.891	.903	.392	.563	.734
Delta 2		.817	.778	.685	.904	.884	.50	.503	.715
Terman Form A			.823	.714	.931	.929	.534	.586	.741
Alpha, Form 8				.712	.842	.914	.471	.564	.716
Mentimeter					.779	.842	.285	.409	.654
Av. 1st 3 Tests						.975	.527	.562	
Av. Five tests above							.453	.60	.841

* An unpublished test of grammar and correct usage arranged by Miss Rewey Belle Inglis, University High School, Minneapolis.

In how many of the 21 cases of wide difference between tests and school marks did further examination show that the first test given was unreliable?

The following are the four pupils whose percentile ranks in the Miller test and in school marks differed by more than 50 points.

Pupil	P. R. in Miller Test	P. R. in Av. of 5 tests	P. R. in School Marks
8	77	82	16
21	60	20	8
14	6	10	70
50	97	100	39

It will be observed that further examination of these pupils with four other tests confirmed their percentile ranks in the Miller Test in 3 out of 4 cases. It is evident that number 21 is not rated properly by the Miller Test. The average of the five tests gives her a percentile rank of 20. One of two explanations is possible: (*a*) previous information about the test, or (*b*) "copying" when the test was given. The former explanation seems the more plausible, since every precaution was taken to prevent the latter. The school marks and the average of five tests place her in the lowest fifth.

It is quite evident that we are not paying dividends on No. 8 and No. 50. Both boys are in the upper 25 percent in ability, but they are distinctly below average in achievement. What is the reason? No completely satisfactory answer can be given at this time, but the following facts make clear the nature of the discrepancy.

Pupil No. 8 made scores on the tests as follows:

Test	Score	P. R.
Miller Mental Ability Test.............	74	77
Haggerty's Delta 2....................	150	88
Terman Test, Form A..................	156	80
Army Alpha, Form 8..................	133	70
Trabue's Mentimeters	121	82
Otis Test	166	65
School Marks (36 weeks)..............	33.5	16

His age is 14 years 2 months. He is very much undersize, undernourished, restless, timid, and somewhat indifferent. His conduct is all that could be desired. He comes from a good home. His father says his son has always been in good health. He has poor study habits. His school work has not improved; P. R. in school marks for first quarter (12 weeks) was 21, second quarter 12, for the year, 16. He presents a clear-cut problem which has not been solved.

Pupil No. 50 made scores and obtained percentile ranks as follows:

Test	Score	P. R.
Miller Mental Ability Test.............	88	97
Haggerty's Delta 2	152	92
Terman Test, Form A..................	173	95
Army Alpha, Form 8..................	166	100
Trabue Mentimeters	130	100
Otis Test	191	98
School Marks (36 weeks)...............	47.9	39

Pupil 50 is 15 years, 2 months of age. He is very much over weight and a "good feeder." He is well behaved, good natured, easily embarrassed, very reticent, and lazy. He is not regular and persistent in his efforts. He has on certain occasions written almost perfect examination papers. He does not conform to class requirements that are necessary to make good marks. He opened the first quarter with a P. R. in school marks of 70 and averaged 39th P. R. for the year. His father is a successful business man. It is clearly evident the school is not getting out of the boy all that he is capable of doing. Why?

Pupil No. 14 shows results quite contrasted to those of No. 8 and No. 50. His record is:

Test	Score	P. R.
Miller Mental Ability Test.............	43	6
Haggerty's Delta 2	117	15
Terman Test, Form A..................	99	18
Army Alpha, Form 8..................	101	13
Trabue Mentimeters	91	5
Otis Test	124	8
School Marks (36 weeks)...............	74.4	70

This boy's age is 13 years, 9 months. He is courteous, industrious, co-operative, and very loquacious. He takes pride in his school work and tries hard to please. He has several interests outside of his school work. He is a slow reader. He is popular with his teachers and classmates, especially with the girls. His home influences are excellent; his father is a professional man. This boy is not a problem for the school. He is, however, a very interesting example of a boy who can make good school marks even though his mental test scores are low.

Below are the results of further examination of the seventeen students whose percentile rank in the Miller Test differed from the percentile rank in school marks from 25 to 50 points.

Pupil	P. R. in Miller Test	P. R. in Av. of Five Tests	P. R. in School Marks
31	84	92	56
36	82	76	36
44	66	58	18

Pupil	P. R. in Miller Test	P. R. in Av. of Five Tests	P. R. in School Marks
45	50	52	16
7	30	30	0
49	60	65	12
27	75	78	100
41	53	72	87
20	9	18	42
4	13	30	61
25	20	42	56
38	35	50	80
23	38	27	83
52	50	47	90
19	30	40	70
39	5	2	46
51	66	70	20

It will be observed that the percentile ranks of the students in the average of the five tests confirm the ratings in the Miller test except in three cases, Nos. 4, 25, and 38, to whom further examination gave percentile ranks from 15 to 27 higher. In all three cases the higher rating is confirmed by the percentile rank in school marks.

In the other fourteen cases we have no reason to believe that the percentile ranks in the tests would be materially modified by giving more than the five tests. The reasons for the difference between percentile rank in tests and in school marks must be attributed to something other than faulty examinations.

Pupil No. 49 is a type well known to most educators:

Test	Score	P. R.
Miller Mental Ability Test.............	68	60
Haggerty's Delta 2	143	73
Terman Test, Form A..................	145	65
Army Alpha, Form 8..................	130	64
Trabue Mentimeters	110	49
Otis Test	166	65
School Marks (36 weeks)..............	28.2	12

He is 14 years old, normal physically. He is a likable boy, with little pride or ambition. He is capable of 'spurts,' but is lacking in sustained effort. Two of his older brothers, more capable than he, have exhibited the same traits in a more marked form. The family is in very good circumstances and both parents are much concerned about the education of their children. During the year the boy made little or no permanent improvement. His next older brother, a sophomore, made no noticeable change for the better during the two years.

Pupil No. 51 is a girl 14 years of age, very much overweight.

Test	Score	P. R.
Miller Mental Ability Test.............	71	66
Haggerty's Delta 2	142	70
Terman Test, Form A....................	135	54
Army Alpha, Form 8.....................	137	81
Trabue Mentimeters	124	92
Otis Test	170	73
School Marks	36.9	20

In early childhood she had spinal trouble which made her an invalid for more than half of her life. Her difficulty seems to be a lack of independence and initiative, due very likely to her experiences as an invalid and an only child. She does what she is told to do and waits for orders. She is gaining in independence. She made considerable progress during the year and will probably continue to improve.

It will be remembered from the explanation of the correlation graph given earlier that the upper left-hand quarter contains the pupils who are in the upper half in the test but in the lower half in school marks, while the lower right-hand quarter contains those who are in the lower half in the tests but in the upper half in school marks. It is interesting to note in this connection that all except one of the seven pupils in the upper left-hand quarter of the graph, Fig. 11, are boys, while all except two of the eight pupils in the lower right-hand quarter are girls.

Furthermore, pupils in the lower right-hand quarter are conscientious, industrious "lesson getters" under parental supervision; but those in the upper left-hand corner cannot be characterized in this manner.

The interesting and important question is whether the pupils in the upper left-hand quarter can be prevailed upon to assume an attitude similar to those in the lower right quarter. When they assume such an attitude, the place they have occupied will be vacant for they will have moved to the upper *right-hand* quarter where they belong.

When the upper left-hand quarter of the graph is densely populated, your school is not paying dividends on the gray matter at its disposal. When you find this condition existing, don't decide too quickly that mental tests are not a measure of mental ability.

Pupils in the lower right-hand quarter, Fig. 11, are in all cases

industrious, courteous, cooperative, dependable, and conscientious. They or their parents, and sometimes both, take pride in school marks and work diligently to get them. They are all good 'lesson getters.' They conform. Without exception they are students with pleasing personalities. Teachers naturally dislike to have them receive low marks. The mental tests don't register these excellent qualities, but the school marks do register them.

Pupils in the upper left-hand quarter are characterized by a different set of adjectives. They are not regular in their work habits. They work by 'spurts' or not at all. They take little pride in their school work, and marks do not appeal to them. They are non-conformists in classroom requirements and are therefore not good 'lesson getters.' Mental tests do not register or measure a pupil's attitude toward a piece of work that requires sustained effort for several hours daily for 36 weeks; school marks are affected materially by such an attitude.

It is rather discouraging to note that very little change, if any, was made in pupils Nos. 49, 50, and 8, who were described above. What change, if any, in the attitude of pupils of this type can be made during four years? Unfortunately, we do not know enough about methods of handling such individuals. A careful record of such cases, including reports of methods of treatment, especially of those methods that bring results in the way of better achievement, would be a great value to all teachers and administrators. A 'case book' including these *types,* for certainly each case is not unique, ought to contribute a great deal to this problem. The problem is an obstinate one. Is it possible that restrictions laid down by physical and social inheritance make it impossible to make desirable changes? Does any one know? What scientific data are available to establish what is possible? We do know that there is a tendency for pupils to retain similar quartile standing thruout the elementary school, the high school, and the college. How many pupils of the types represented by Nos. 8, 49, 50 and 51, or what proportion of them, never do a quality of work in keeping with their mental ability? Can this proportion be reduced, and if so, by what methods?

In the opinion of the author one of the chief benefits to be derived from mental testing is the direction of the attention of teachers and principals to individual pupils of the 'could-if-they-would' type. This benefit can be realized whether or not the pupils are classified; however, classification on the basis of mental ability will place the pupils in an environment better adapted to their needs and capacities.

In this discussion emphasis has purposely been placed on those cases with which the University High School has failed, in order to set forth more clearly the problem involved.

CHAPTER VIII

SOME ADMINISTRATIVE USES OF INTELLIGENCE TESTS IN THE NORMAL SCHOOL

Bessie Lee Gambrill
Head of the Department of Psychology, New Jersey State Normal School, Trenton, New Jersey

In the very brief time available for preparing this report it was impossible to attempt any general survey of the administrative uses of tests in the normal schools of the country. All that seemed feasible was to make a report of three years' experience with the Thorndike Intelligence Examination for High-School Graduates in the normal school with which the writer is connected, and to supplement this report by such data as could quickly be gathered from some normal schools that the writer knew had given intelligence tests.

I. Intelligence Tests at Trenton

The New Jersey State Normal School at Trenton has been using the Thorndike Intelligence Examination since the fall of 1919. During this time investigation has been directed chiefly toward the discovery and testing of the possible administrative uses of the test. It was hoped especially that such a test might ultimately provide a sound basis for sectioning students according to intellectual ability, furnish a check on the teacher's judgment of ability, help to identify early the student who lacked the ability to complete a normal-school course and the student who was able but who would not work.

The first test was given in the fall of 1919 to the entering (junior) class. An attempt had already been made to group this class according to scholastic ability. Since no other measure was as yet available, high-school marks had been made the basis of sectioning. At the end of the first semester, therefore, three independent means of ranking these juniors were available: first, the high-school marks; second, the test scores achieved; and third, the

teachers' first semester marks, since the faculty had been told nothing, until after these marks had been reported, either as to the order in which the sections were ranked or as to the test scores achieved by the students. It was desirable to know the extent of agreement among these three independent measures.

The first question considered was, how far the sectioning according to scholastic ability would have been altered if intelligence scores rather than high-school marks had been the basis of grouping. To furnish an answer to this question each section was charted in such a way as to show the number of individuals whom the intelligence scores would displace from the sections to which they were assigned on the basis of high-school marks, and the degree of such displacement in terms of sections. Only general course students could be included in this study, since students taking special courses—Domestic Science, Kindergarten-Primary, Music, etc.—had been sectioned according to the special interests and not according to the high-school marks. A commuter's division, which was not grouped on the basis of marks, also had to be omitted. These omissions left four sections ranked according to high-school marks. In the following tabulation these sections will be designated A, B, C, and D; A is the highest ranking section and D the lowest ranking section. Table I shows the extent to which this sectioning would have been altered, had it been determined by the intelligence scores.

Table I shows that 36 of the 95 students were not displaced from their sections by the test. That is to say, in 38 percent of

TABLE I.—DISPLACEMENT OF STUDENTS BY THORNDIKE INTELLIGENCE EXAMINATION SCORES FROM SECTIONS TO WHICH THEY WERE ASSIGNED ON THE BASIS OF HIGH-SCHOOL MARKS

Amount and Direction of Displacement	Section A	Section B	Section C	Section D	Totals
+3	0	0	0	3	3
+2	0	0	6	6	12
+1	0	8	1	8	17
0	14	7	9	6	36
—1	6	5	8	0	19
—2	2	3	0	0	5
—3	3	0	0	0	3
Totals	25	23	24	23	95

the cases considered, there was perfect agreement between the high-school marks and the intelligence test in sectioning students according to intellectual ability. If we add to the 36 individuals with zero displacement the 36 whom the test would have pushed up or down but one section, we find that in approximately 76 percent of the cases the two methods of sectioning do not disagree by more than one section. Six individuals, or six percent of the group, however, would have been exactly reversed as to section had they been assigned on the basis of their test scores. Three students in Section A, the highest ranking section, would have been in Section D, the lowest ranking section, and three who were in Section D would have been in Section A.

Since the purpose of the sectioning is to group together those students who can progress in school work at approximately the same rate, it was important to know whether the high-school marks or the tests were more accurate in placing together students who succeeded in the accomplishment of normal-school work to approximately the same degree. The second question considered, therefore, was the sectional displacement which would occur should the students be regrouped on the basis of the teachers' marks for the first semester's work in the normal school. For the purpose of answering this question the groups sectioned according to high-school marks were recharted so as to show the displacement which teachers' marks would occasion. Table II, which presents the results,

TABLE II.—DISPLACEMENT OF STUDENTS BY FIRST-SEMESTER NORMAL-SCHOOL MARKS, FROM SECTIONS TO WHICH THEY WERE ASSIGNED ON THE BASIS OF HIGH-SCHOOL MARKS

Amount and Direction of Displacement	Section A	Section B	Section C	Section D	Totals
+3	0	0	0	3	3
+2	0	0	3	7	10
+1	0	7	7	3	17
0	11	9	7	10	37
—1	6	5	7	0	18
—2	5	2	0	0	7
—3	3	0	0	0	3
Totals	25	23	24	23	95

reveals the fact that 39 percent of the 95 students are not displaced by the first-semester normal-school marks from the sections to which they were assigned on the basis of high-school marks, that 75 percent are not displaced by more than one section, and that 6 percent are displaced from the lowest to the highest, or from the highest to the lowest section. These percentages are in striking agreement with the percentages representing the correspondence between high-school marks and the Thorndike Intelligence Examinations as bases of sectioning. Analysis of the original chart, however, showed that the two measures, marks and tests, did not agree quite so perfectly as to the individuals displaced. It did show, however, that there was less discrepancy between the test scores and the normal-school marks than between the high-school marks and either test scores or normal-school marks.

The results secured from this first test convinced the faculty that the test gave promise of serving valuable administrative ends. All conclusions formed, however, were tentative, and needed to be verified by further study. It was seen, for example, that if the intelligence test could locate those individuals who had not the ability to complete the normal-school course, many students might, through a three-hour examination, be spared the time, expense, and humiliation of spending from half a year to a year and a half in the normal school only to discover finally that they could not be graduated. To locate the limits within which students must test in order to have a reasonable hope of graduation would require careful study, for several years, of the scholastic careers of students in relation to their test scores.

As a direct measure of the probable relationship between the first-semester normal-school marks and the Thorndike test scores, the coefficient of correlation between the two measures was computed.[1] The correlation calculated by the 'foot-rule' formula, was .56, P. E. .03.

The correlation between the Thorndike Intelligence Examination and first-semester college marks for 500 freshmen in Brown

[1] The Trenton Normal School uses a five-point scale of marking: A, B, C, D, F. To obtain a student's scholarship mark for correlation, the marks assigned him were translated into arbitrary numerical equivalents (A, 7; B, 5; C, 4; D, 3; F, 1) and averaged.

University, Columbia College, and Rutgers College is reported by Thorndike as about .55. Thorndike says of this correlation: "When allowance is made for 'attenuation' of the correlation by the lack of precision in a rating on only one half year's work, this will rise to .60 or more. . . . Since college achievement is in part due to factors of health, ambition, economic conditions and the like, the correlation between the Thorndike score and the intellectual factors of college achievement alone may be put somewhere between .85 and .95 for a group of high-school graduates in general." There seems no reason to doubt that these facts would hold for normal-school students as well as for college students so far as the academic side of the normal-school course is concerned.

These conclusions were borne in mind in the study of individual cases which followed. A comparison of the score achieved by the student with his actual class accomplishment revealed in certain cases the fact that he was not working up to his capacity. The causes for the discrepancy were then sought. In some cases these were found to be physical difficulties; in others, poor health habits, timidity, wrong attitude, poor habits of work, outside distractions or laziness. The test gave the teacher confidence that, in applying the spurs to the student with a high test score and poor scholarship, he was not demanding the impossible. In the case of students with low scores and records that were low, but not low enough for failure, patience was the only reasonable course, since they were doing as well as their endowment permitted them to do. For the remainder of the year the intelligence records were consulted whenever a teacher was in doubt as to whether a student was measuring up to the scholastic standard of which he was capable. While no student was dropped from the school because of a low test score, it is safe to say that since the first use of the test no student has been dropped from the school for poor work, without consideration of his rating on the intelligence test.

In the fall of 1920, this class was retested—in part, to measure the reliability of a score based on a single performance, and in part to make clear the meaning of the test by furnishing an answer to the following question which had arisen: "Will a re-test measure a student's improvement in ability from a year's work

in the normal school?" Unfortunately, one section of the class, the strongest section, did not take the re-test because its members were doing their practice teaching. For those students (169 in number) who took both tests, the coefficient of correlation between the scores they attained as juniors and the scores they attained as seniors was .86, P. E. .01 (Pearson formula); that is, the agreement between the two tests was close, but as might be expected, not perfect. Differences between the two were, in general, small. In a few cases, however, they were large enough to emphasize the danger of taking any decisive action, such as the exclusion of a student from school, on the basis of a single test, unless the test was supplemented by other measures of his ability.

There was no consistent tendency for the re-test scores to be better than the original scores. About 60 percent, however, made somewhat better scores on the re-test. Since the differences were, in general, very small, the slightly greater tendency to do better on the second test was probably due, in part at least, to the fact that the situation had ceased to be entirely new. Certainly the re-test showed nothing to indicate that it could serve to test improvement gained from the year's work in the normal school.

Study of the results of the tests given to the junior classes entering in 1920 and 1921 has served to confirm the judgment that the Thorndike Intelligence Test scores give a reasonably reliable basis for predicting a student's ability to meet the scholastic demands of the normal-school course. In 1920 the instructors were asked to hand to the Psychology Department a list of the poorest tenth of their juniors. This list was prepared before the results of the intelligence tests were reported to the faculty. Upon tabulating the returns it was found that the five students who scored below thirty had been reported as unsatisfactory by a majority of the teachers to whom they recited and that a majority of those who scored below 40 had been reported as unsatisfactory by two or more of their instructors. In December, 1921, that is, a year and a half after they entered, a tabulation was made showing the status of these students who tested below 40, with the following result:

STUDENTS WHO ENTERED SEPTEMBER, 1920, AND WHO TESTED BELOW FORTY IN THE THORNDIKE INTELLIGENCE EXAMINATION

Score	*Status, December, 1921*
21.4	Withdrew because of unsatisfactory work
22.3	Withdrew because of unsatisfactory work
24.0	Withdrew because of unsatisfactory work
25.1	Withdrew because of unsatisfactory work
30.0	Withdrew because of unsatisfactory work
30.2	Withdrew because of mother's death
30.5	Advised to withdraw
30.5	Withdrew
33.2	Low, but passing record; hard worker
34.2	(Domestic Science) Marks vary from A to D
34.4	Must extend course one-half year
34.6	Withdrew
34.7	Must extend course
34.7	Variable record
35.0	Withdrew
36.8	Must extend time
37.0	Withdrew
37.6	Must extend course
37.6	Must extend course
38.4	Withdrew
38.6	Withdrew
38.8	Withdrew
38.8	(Domestic Science) Marks from A to D
38.9	Must extend course
39.4	Poor record. Many F's and D's
39.5	Must extend course
39.9	Must extend course

The majority of withdrawals occurred as the result of advice or pressure from the school, or as a result of the student's own realization that he lacked the ability to meet the school's requirements.

On the basis of such records as these, the following tentative conclusions seem justified:[2]

First, it is highly probable that any high-school graduate testing below thirty on the Thorndike scale lacks the intellectual ability necessary to complete the course in this Normal School. The available data include the scores of the class of 1920,[3] the class of 1921 and the class of 1922. No student with a score of thirty or below has been graduated, and, as indicated in the foregoing tabulation,

[2] Any conclusions as to the value of intelligence tests are based on the assumption that the tests were carefully given and scored under the direction of a competent person familiar with the requirements for scientific testing.

[3] Tested in June of the senior year. The tests were scored by Mr. F. L. Whitney of the University of Minnesota, who is using the results in a study of intelligence tests in relation to success in teaching.

all students in the class of 1922 testing thirty or below, have already (December, 1921) been eliminated.

Second, a majority of the pupils testing between thirty and forty will probably not complete the course, or will do so only by remaining in the normal school for an extra half year or longer. Whether or not the school is justified in retaining these students who can complete the course only by taking longer than the allotted time, can only be determined by watching the careers of this experimental group.

Study of the distribution of test scores for all classes examined revealed a number of interesting facts. Table III shows the distribution of scores attained by four successive June classes, and by three February classes. The year designated is the year of graduation. The February classes were tested the fall after they entered. The class of 1920 was tested a few weeks before graduation. The other three classes were tested at the beginning of their junior year.

The distribution of scores in Table III reveals the intelligence level of students entering the normal school and makes possible a comparison between the intellectual caliber of these students and of students entering the freshman class in certain colleges. The scores attained by the classes which entered the Trenton Normal School in September 1919, September 1920, and September 1921, were compared with the scores attained by two groups of women college students; (1) "Freshmen, Liberal Arts College, Eastern State," and (2) "Freshmen, Home Economics, Western State." The distribution of scores for these women is given by Thorndike in his summary on the "Significance of Scores in the Thorndike Intelligence Examination for High School Graduates." The comparison shows that the Liberal Arts college draws a much larger proportion of high-ranking students than does the normal school. Only 15 percent of the normal-school students reach or surpass the median for this group of college women. The normal school suffers little, if any, however, by comparison with the Home Economics women. Table IV shows comparatively the distribution of scores for these three groups. The figures are only approximate.

TABLE III.—THORNDIKE INTELLIGENCE EXAMINATION FOR HIGH-SCHOOL GRADUATES
NEW JERSEY STATE NORMAL SCHOOL AT TRENTON
Comparative Distribution of Scores for Four Successive Classes

Range of Scores	Percentage of Class Achieving Specified Scores						
	Class of 1920	Class of 1921	Class of 1922	Class of 1923	February Classes 1921	1922	1923
20.1– 30	0.	0.5	1.8	0.6	0.	0.	0.
30.1– 40	3.	1.5	8.5	6.9	0.	0.	0.
40.1– 50	16.	12.7	21.4	19.2	7.1	36.8	25.0
50.1– 60	21.	23.4	29.1	31.5	21.4	10.5	25.0
60.1– 70	31.	28.3	23.7	27.4	42.8	31.6	25.0
70.1– 80	17.	20.5	11.9	11.0	7.1	10.5	16.6
80.1– 90	10.	9.6	1.1	3.2	21.4	10.5	8.4
90.1–100	2.	3.9	0.0	0.0	0.	0.	0.
100.1–110	0.	0.5	0.0	0.0	0.	0.	0.
Median	62.6	64.1	55.9	57.8	64.4	60.8	59.8
Lowest Score	34.2	24.0	21.4	27.1	48.5	41.2	45.9
Highest Score	97.9	100.3	83.0	84.9	87.9	81.8	83.9
Range of middle 50 percent	55.3–72.5	56.0–74.2	47.7–64.4	49.5–65.	53.3–78.3	46.2–71.6	50. –67.4
Number	100	207	269	317	14	19	12
Test Forms Used	Part I, E and F Part II, C Part III, C	Part I, C and D Part II, B Part III, B	Part I, I and M Part II, D Part III, D	Part I, I and M Part II, D Part III, D	Part I, C and D Part II, B Part III, B	Part I, I and M Part II, D Part III, D	Part I, I and M Part II, D Part III, D

TABLE IV.—PERCENTAGE OF FIRST-YEAR NORMAL-SCHOOL STUDENTS AND COLLEGE FRESHMEN ATTAINING CERTAIN SCORES ON THE THORNDIKE INTELLIGENCE EXAMINATION

Score	Freshmen Liberal Arts Eastern State	Freshmen Home Economics	First-Year Trenton Normal (3 classes)
100	0	0	0.2
90	8	1	1.2
80	28	7	5.5
70	58	24	19.3
60	86	45	45.7
50	94	77	74.3
40	98	96	92.6
30	100	100	98.6
20	—	—	100.0
Approximate Median	72	58	58.8

By applying to the normal-school group the Thorndike standards for prophesying college success on the basis of intelligence scores, a comparison was made between the intellectual ability of the normal-school students and the ability required for successful college work. Thorndike's interpretation of scores for a high-grade college follows:

A boy scoring over 95 is worth admitting in almost entire disregard of technical deficiencies.

A boy scoring 85 to 95 has intellect enough to do collegiate and professional work with distinction.

A boy scoring 70 to 85 has intellect enough to do the work to obtain a college degree.

A boy scoring 60 to 70 may be admitted if he is sufficiently in earnest and otherwise desirable.

A boy scoring 50 to 60 should be admitted only if he is of extraordinary zeal or has suffered very great educational handicaps.

A boy scoring under 50 should not be admitted.

He suggests that since the test "perhaps slightly penalizes girls in comparison with boys, having been designed primarily for the latter," present standards may be set five points lower for girls than for boys. Since the overwhelming majority of Trenton students are girls, this adjustment of standards was made. The following summary shows for the normal-school group the prophecy of success in intellectual work of the quality demanded for gradua-

tion from a high-grade college, in terms of the modified Thorndike standard:

Approximately 1.5 percent score over 90. These might be admitted to work of collegiate grade in almost entire disregard of technical deficiencies.
Approximately 4 percent score from 80 to 90. They could do collegiate and professional work with distinction.
Approximately 26 percent score from 65 to 80. They have intellect enough to do the work to obtain a college degree.
Approximately 28 percent score from 55 to 65. They might be admitted if sufficiently in earnest and otherwise desirable.
Approximately 24 percent score from 45 to 55. They should be admitted only if they possess extraordinary zeal or have suffered very great educational handicaps.
Approximately 16 percent scored below 45. These students should not be admitted to work of college grade.

So far as this group of students is concerned, then, 6 percent are capable of doing work of college grade with distinction; an additional 26 percent have sufficient intellect to do successfully the work necessary to win a degree; 50 percent might be admitted to work of college grade only under very special conditions; 16 percent test so low that they should not be admitted to work of college grade under any circumstances.

Such an analysis and comparison of intelligence levels is administratively important as a basis for considering modifications in curriculum and method, and as a basis of adjusting with colleges and universities the amount of credit to be allowed for normal-school work. Also the wide variation in the intellectual abilities of normal-school students which is thus thrown into relief, re-emphasizes the necessity of giving due weight to the matter of intellectual ability in sectioning students for purposes of instruction. The attempt to teach in the same classes, students who are capable of doing college work with distinction and students who are intellectually incapable of doing such work at all, must inevitably be unprofitable and wasteful, if not wholly disastrous to one or both types of student.

Inspection of Table III not only reveals wide variations in individual ability but also shows differences in intellectual ability

in classes entering in different years. Disregarding the class of 1920, which was tested at the end of the senior year, and the February classes, which were tested six months after entrance to the normal school and so presumably had eliminated by this time their weakest students, there is a marked contrast in the distribution of scores for the class of 1921 and the two classes which follow it. The probable explanation of the intellectual superiority of the class of 1921 is found in the fact that it entered the normal school at a time when economic motives urged earning rather than studying and business rather than teaching. The normal school consequently drew a smaller and more highly selected group than it has drawn in the two succeeding years. During these two years every possible appeal has been made to induce high-school seniors to prepare themselves for teaching. No corresponding care has been taken to measure the mental status of those who have responded to the appeal.

In addition to the general course, which qualifies for teaching in any grade of the elementary school, Trenton offers a number of special courses: a kindergarten-primary course, which prepares for teaching in the first four grades; a domestic science course, a commercial teacher's course (3 years), a music supervisor's course (3 years), a manual training course and a physical training course. An analysis of the intellectual level of the student body should include a study of the test scores of students electing these special courses, comparing each course with the other special courses, and with the general course, as to its intellectual level. Table V enables us to make such a comparison for two classes.

The data given in this table serve to define an important administrative problem whose solution demands still more data of the same type, a careful study through a series of years of the educational and professional careers of these special students and an analysis of the abilities which their special work demands. Graduates of the Physical Training group and of the Music group will be called upon not only to teach children but also to supervise the work of other teachers. From this point of view, theoretically a higher type of intelligence should be demanded for acceptance of candidates for these special courses. Do other factors in the situa-

TABLE V.—DISTRIBUTION OF THORNDIKE INTELLIGENCE SCORES FOR JUNIORS IN DIFFERENT COURSES OF THE NORMAL SCHOOL AND FOR DIFFERENT DIVISIONS OF THE GENERAL COURSE, SECTIONED ACCORDING TO HIGH-SCHOOL MARKS

Course	Median		Highest Score		Lowest Score		Limits of middle fifty percent		Number of Cases	
	Class		Class		Class		Class		Class	
	1922	1923	1922	1923	1922	1923	1922	1923	1922	1923
Commercial (3 years)	54.2	57.0	67.4	74.7	38.8	31.4	61.0–49.4	41.1–63.6	11	19
Domestic Science	56.8	53.7	80.4	83.5	22.3	30.8	53.0–70.0	41.9–60.2	20	13
Kindergarten-Primary	51.0	56.4	76.5	78.5	24.0	31.2	45.2–60.1	45.9–69.1	43	48
Manual Training	—	57.3	—	59.6	—	54.6	—	54.9–59.5	—	4
Music (3 years)	56.5	59.0	68.0	81.7	37.0	42.3	48.0–61.0	49.0–66.1	6	10
Physical Training	56.8	58.3	71.8	79.2	44.9	45.9	50.0–62.6	53.7–65.9	18	26
General Course—										
Commuters	55.4	61.4	76.6	80.6	25.1	38.2	40.2–64.4	48.4–66.0	22	20
Section A*	66.1	65.7	83.0	83.4	40.0	34.1	60.8–69.8	75.5–59.4	25	30
Section B	63.9	61.0	78.6	83.0	47.1	36.8	57.3–71.0	50.2–67.5	26	30
Section C	54.8	57.9	82.4	84.9	37.6	28.2	45.3–64.4	53.4–67.0	28	29
Section D	51.8	56.8	77.5	78.1	33.2	45.0	44.7–63.7	49.7–65.5	22	29
Section E	51.1	55.5	70.8	83.3	30.5	27.1	45.9–60.0	49.7–62.4	22	30
Section F	50.2	53.5	69.7	76.1	21.4	30.6	42.6–57.8	47.3–60.5	25	29
Entire Class	55.9	57.8	83.0	84.9	21.4	27.1	47.7–64.4	49.5–65.0	269	317

* A is the highest ranking section.

tion make this demand unwise? Table V shows that the medians for these two sections are somewhat higher than the medians for the class as a whole in both years for which data are presented.

In general, the table shows no conspicuous tendency for the special-course students to test on the average higher or lower than the general-course students. The kindergarten-primary group, class of 1922, and the domestic science group, class of 1923, do test markedly lower as groups than do the classes of which they are a part. It is administratively important to consider whether the conditions required for success in the fields for which these courses prepare, demand changes in the selection of students for these courses.

Table V also shows the medians, highest, and lowest scores and the ranges of the middle fifty percent of students in the general course. The measures for the commuters' section parallel very closely the measures for the class as a whole. The remaining six sections, grouped according to ability on the basis of high-school marks, show by their medians that an attempt to section according to ability even on this basis does produce a somewhat more homogeneous grouping than a hit-or-miss procedure. Comparison of the range of scores, however, and of the limits of the middle fifty percent, indicate the necessity of re-sectioning if anything like homogeneous groups are sought, and this re-sectioning will be done at the beginning of the second semester.

In a professional school for teachers it is important to discover early, not only a student's scholastic promise, but also the probability of his success in his actual work as a teacher. To what extent can the student's intelligence score be taken as a prophecy of his probable success in practice teaching and of his success in classroom teaching after graduation? The only objective evidence that can be offered from the Trenton Normal School at this time is a correlation between the practice teaching marks and the intelligence scores of the class of 1921. This correlation, calculated by the Pearson product-moment formula, is .11; P. E., .05. If other data, which will soon be available, should support this evidence of the low relationship between the intelligence score and success in classroom teaching, it will be highly important for normal schools to investigate every method of measurement that offers hope of

discovering and testing the abilities other than abstract intelligence, required for success in teaching. Trenton expects to give the Downey Will-Temperament test in the near future and to study the results in relation to classroom success. The Millersville Normal, Pennsylvania, is also planning to study the possible value of this test.

While the Trenton Normal School will maintain the experimental attitude toward its use of intelligence tests—attempting to analyze its results more fully, checking its tentative conclusions by further study, supplementing from time to time the test now in use by such others as may offer hope of throwing light on the more effective conduct of teacher training, no doubt remains as to whether an intelligence test is a valuable administrative tool. Such a test has become a necessity.

Experimentation with the Thorndike Intelligence Examination in this school seems to justify the following summary of administrative uses, actual or potential, of such a test in normal schools.

1. The test is valuable, and should yearly become more valuable, in helping to locate (*a*) students who have not sufficient intelligence to complete a normal-school course, (*b*) students who have sufficient intelligence to complete the course only if given more than the allotted time, (*c*) students who are capable but who make poor grades because they are lazy, physically unfit or have temperamental defects which interfere with scholastic success.

2. The test furnishes a valuable basis for conference with students who are doing poor work or who are doing work of a quality poorer than their ability warrants. The dean, student advisor or teacher will find the intelligence test score a welcome check on his own personal judgment of the student's mental ability.

3. The test scores provide an objective basis for sectioning students according to their intellectual ability.

4. The intelligence records provide a valuable basis for conference with high-school principals with respect to the quality of work done in the normal school by their graduates.

5. The records provide an argument for the administration of intelligence tests in high schools and the consideration of scores

there achieved as one basis for advising students as to the wisdom of entering the normal school.

6. The most far-reaching potential administrative use of the test is that it may serve as a research tool of the greatest ultimate value in helping to analyze and define the problems of teacher training. Evaluation of curricula and methods can proceed scientifically only in the light of knowledge of the human material to which they are to be applied. Analysis of the raw material of teacher training is logically the first step toward determining the most effective handling of this material and toward trying to secure for the future a higher average of recruits for the teaching profession.

The experience with the tests also suggests certain cautions that should qualify the administrative uses of intelligence tests.

1. The tests should be given and scored under the direction of a competent person who is familiar with the requirement for valid testing. Record should be made of any unusual condition prevailing at the time of testing. A low score made by a strong student was explained by an examiner's note that Mr. X was evidently suffering from a severe cold. A high record made by a poor student was understandable in the light of an examiner's note that Miss Y copied from a neighbor.

2. No radical action, such as advising a student to withdraw from school, should be based upon the results of a single test, unless the conclusion from the score is supported by other measures of ability, such as high-school marks or teachers' judgments. Provision should be made for additional tests in doubtful cases.

3. Intelligence tests will not give all the facts that are required for prognosis of a student's probable success as a teacher. While there is unquestionably an intelligence level below which no one could fall and still succeed as a teacher, that point can be determined only tentatively at present. Somewhere along the line there may be a point above which additional increments of "intelligence" do not bring increased potentialities for success as a teacher. Certainly there are other qualities, the absence of which will cause failure in teaching no matter how highly endowed intellectually the individual may be. Experience shows that high test scores

alone do not insure success in practice teaching or in teaching after graduation. This fact, however, does not destroy the value of the intelligence test. It indicates, rather, the need of supplementing this test by other means of measurement. If reliable tests of temperament, executive ability, and the like can be developed, they will be of inestimable value. The writer believes that in the meantime, high schools and normal schools should keep records of the extra-curricular interests and activities of their students, and study the possible significance of these records in relation to qualities other than abstract intelligence, which may condition success in teaching.

Intelligence Tests in Certain Other Normal Schools

Prior to the current year a number of Pennsylvania Normal Schools had given the Thurstone Test IV Psychological Examination. The writer secured no report, however of any administrative purposes to which this test may have been put. Two normal schools, Slippery Rock and Millersville, in 1920-21 gave Trabue's Mentimeter, School Group 2A. In the *Pennsylvania School Journal* for October 1921, Mr. J. B. Thomas, head of the department of Education at Millersville, describes the results of this test. The interesting feature of the report from the standpoint of possible administrative uses of such a test is a comparison of the median scores attained by students electing different curricula in the normal school. Curriculum I is elected by students who are to teach in grades one to three; Curriculum II by those who expect to teach in grades four to six; Curriculum III, by those who will teach in grades seven to nine or in the junior high school; Curriculum IV by those who will teach in rural schools. Mr. Thomas reports these results:

Median score of all Juniors	119.5
Median score for Curriculum II	108.5
Median score for Curriculum I and IV	117.5
Median score for Curriculum III	126.5

During the current year the Bureau of Teacher Training of the Pennsylvania State Department of Education has directed the giving of an intelligence test in all Pennsylvania Normal Schools. The

test used was a part of the Thorndike Intelligence Examination for High-School Graduates, Part I, forms I and M.[4]

The data for presenting comparative results of the Thorndike tests in the different Pennsylvania normal schools were not available in time for inclusion in this report. Such records as were available showed no marked variations in the intellectual quality of students in different normal schools. The medians, the highest and lowest scores, and the score limits of the middle fifty percent of students in the Pennsylvania normals also indicated that their intelligence level was approximately the same as that of the students in the Trenton, New Jersey, Normal School.

One table of results, furnished by the Indiana (Pennsylvania) Normal School, is reproduced here because it furnishes another comparison of the intelligence levels of students electing different courses in the normal school.

TABLE VI.—SCORES FOR THORNDIKE INTELLIGENCE TESTS; INDIANA STATE NORMAL SCHOOL, INDIANA, PENNSYLVANIA

Group	No. of Students	Highest Score	Lowest Score	Median Score	Range of Middle 50 percent
All Regular Seniors	211	276	110	196	170–216
Regular Seniors; Junior-High-School Curriculum	49	276	132	211	195–234
Regular Seniors; Intermediate Curriculum	73	253	110	195	171–216
Regular Seniors; Primary Curriculum	87	258	110	191	168–207
Regular Juniors	214	262	63	183	162–203
Special Art Students	6	242	152	196	183–225
First-Year Commercial	54	247	112	184	166–197
Senior Commercial	25	229	117	189	164–219
First-Year Home Economics	20	216	102	183	156–199
Senior Home Economics	21	215	104	158	140–193
First-Year Music	11	229	150	176	163–213
Senior Music	12	215	140	177	160–190

[4] It may be of interest to note here that the correlation between Part I, I & M scores, and the total score for the Thorndike examination, computed for a class of 205 juniors at Trenton is .87. The correlation between the total score and first semester marks is .55 and the correlation for the same individuals between the sum of Part I scores and first semester marks is .45.

Actual administrative uses of tests were reported by Pennsylvania Normal Schools as follows:

1. The tests are used by teachers or by the Dean in dealing with individual students in Mansfield, Millersville, Shippensburg, Manchester, and Slippery Rock.

2. Test scores are used in conferences with parents at Mansfield.

3. The test score is made a part of the personal record of the student and is taken account of in making recommendations for positions by Millersville and by Slippery Rock.

4. The test score is a factor in determining whether a student shall "pass" at Millersville. More is demanded from capable students in order to pass.

As possible additional uses, Mansfield and Clarion suggest that tests might be valuable in guiding students in the selection of subjects and in the election of the curriculum to be followed. Slippery Rock ventures the *hope* that the use of the intelligence test may eventually result in the elimination of those who very plainly have not the intelligence necessary to make successful teachers.

Dr. Rowland, Director of the State Bureau of Teacher Training, Pennsylvania, says that the department plans to use the test results in the following ways:

"First, for a comparative study of intelligence levels of our normal-school students with established standards.

Second, for a comparative study of the intelligence levels of the students in the several Pennsylvania normal schools.

Third, for a comparative study of intelligence levels of students in successive years.

Fourth, for a determination of the correlation between these intelligence levels and

a. Results of physical examinations.
b. Social and economic background.
c. Secondary education record.
d. Type of secondary school attended.
e. Normal-school group elections (kindergarten-primary group, intermediate group, junior-high-school group, rural group).
f. Normal-school scholastic record.
g. Normal-school practice teaching record.

The Connecticut State Normal School at New Britain gave the Thorndike Intelligence Test to its entering class this fall. The principal writes:

"I am hoping that certain results will be attained. First, they will give us a basis for conferences with high-school principals concerning the character and attainments of the pupils they send us. Second, they will enable us to compare the general quality of pupils entering the normal schools with freshmen in colleges, and if our standards are too low we may bring pressure to bear to have them raised. Third, I hope the tests may make it possible for the teachers of the school to have a better acquaintance with their pupils."

Work with intelligence tests at the Maryland State Normal School, Towson, Maryland, is reported by J. L. Dunkle and Nellie W. Birdsong of that institution as follows:

We have had three definite aims in mind in the use of various tests with entrance classes: first, to set up equal-ability groups; second, to enable instructors to know better the several abilities of their classes and thus adjust subject matter and method of these; and third, to forecast the probable success of students, and to *check* on outstanding cases that are not measuring up to their tested ability.

In September, 1920, by the Otis Group Test, the entrance class of 120 students was grouped into three sections. The correlations between intelligence scores and academic standing for the year were: Section I, .21; Section II, .26; Section III, .38.

In September, 1921, the entrance class of 280 students was given the Thorndike-McCall Reading Test, and from the data secured they were grouped into six sections. Later the Terman Group Test was used to check the reliability of the grouping. The students could not be reclassified on the basis of the Terman Test because of schedule difficulties. At the end of twelve weeks, the first term, correlations by sections were made between the Terman rating and academic ranks, with the following results: Section I, .67; Section II, .50; Section III, .47; Section IV, .42; Section V, .37; Section VI, .53.

The low correlations between the Otis Group Test and academic rank may be due to any one of three factors or any combination of these, viz: (1) A certain antagonism between equal ability groupings and our marking system; (2) Failure of a *single* test to give or make possible homogeneous groupings; (3) Overconscientious tutelage on the part of the instructors of the weaker student groups.

The higher correlations of the Terman Group Test and academic rank may be explained as follows: (1) The Terman Test is better adapted to the age and status of our students than is the Otis Test: (2) A certain antagonism between equal ability groupings and our marking system may apply here but

will disappear as a factor for consideration when instructors are skillful in using and interpreting a grading system by letters.

In our opinion correlations between the Thorndike-McCall and the Terman Test show conclusively that the former can not be used very helpfully to group students according to ability.

Faculty opinion may be summarized thus: that the Terman Group Test clarifies the instructor's problem by giving her a chance to adapt method and subject matter to the normal, supernormal and subnormal groups; that the mental test helps her to stimulate the individual student to the realization of his possibilities and to keep him working toward that realization.

We have reached one conclusion, and it is that the school should provide educational guidance for those students whose repeated failures or extremely poor work and mental rating are in agreement. The ultimate result may be to direct such students into other fields.

The extension of administrative uses of intelligence tests in normal schools and the assurance with which administrative action may be based upon test results are dependent, in the writer's opinion, upon the building up of standards and the interpretation of results which will follow the bringing together and comparison of experience by all the normal schools which have been experimenting along these lines. It is hoped that the National Society for the Study of Education or some other national organization may in the near future make possible the assembling and presentation of this collective experience.

CHAPTER IX

THE USE OF PSYCHOLOGICAL TESTS IN THE ADMINISTRATION OF COLLEGES OF LIBERAL ARTS FOR WOMEN

AGNES L. ROGERS
Professor of Education, Goucher College, Baltimore, Maryland

At one time in their history there was little danger of the Women's Colleges of Liberal Arts receiving students who were unlikely to benefit by a higher education. Women who sought college training were in general of high intellect and character. The road to college in those days, however, had to be stormed by women, whereas at the present time it is an open highway. Thus candidates for admission have greatly increased in number and represent a more varied sample of interests and abilities than in the past. It is most improbable that only the industrious, the studious, and the intellectually gifted now apply for entrance. The women's colleges are therefore faced with the same problem of selecting their student body as the corresponding institutions for men. Lacking the capacity to provide for the vast numbers clamoring for a college education, they must perforce carefully evaluate their methods of admission with a view to maintaining only those which can lay claim to being sound and right. Not only is it undesirable that they should invest money in training women who are unlikely to profit by advanced instruction, but it would also seem unfair in a democracy to accept the less gifted among women, while those more richly endowed were unprovided for.

Psychological tests form one solution of this problem, which is now being carefully evaluated. Mental tests have, of course, been applied very generally in the women's colleges. They have varied greatly in nature in accordance with the interests of the psychologist in charge and as a rule the abilities measured have been investigated for their own sake rather than for any help they might lend to the

administration of the institution. Tests of color vision, for example, were made at Mount Holyoke over a period of years. At Vassar College the desirability of mental tests as an aid in the forecasting of academic success was early realized and experimentation with a variety of these has been carried on for several years.

The successful application of group tests on a large scale by the United States Army revealed in unmistakable fashion their value as a means of selection and classification on the basis of general ability. This led Goucher College in 1918 to investigate the reliability of those tests which seemed best adapted to differentiate between higher levels of intelligence, with a view to determining their merits as one element in the machinery of admission and also as an instrument for the classification of students in the large required courses. For this purpose use was made of the Thorndike test of Mental Alertness in 1918, supplemented by other tests, and of the Thorndike Intelligence Examination for High School Seniors in 1919 and 1920, and of the Thurstone Psychological Examination for College Freshmen in 1920.

It has already been demonstrated that these tests have much value for these purposes. It has been shown, for instance, that they foretell achievement in the freshmen year with greater accuracy than the previous school record. Again, it has been found that the correlation between the test results and collegiate work in the first year is notably higher than between the ordinary types of entrance examinations and freshmen grades. In general, the latter amounts to less than .45, whereas the coefficient found between psychological test scores (Thorndike Intelligence Examination) and freshmen academic grades has in the case of Goucher College students reached well over 0.60. The prognostic value of the tests is therefore highly satisfactory. They are of undoubted service as an additional check on other data determining fitness for admission.

Their utility in maintaining a high level of student body is not limited to aiding in the selection of students for entrance. They can be an important factor in settling cases of elimination from college. For example, a student of superior intelligence may possibly carry college work with moderate exertion of effort; but stu-

dents in the lowest ten percent of college women in ability can never hope to cope with academic subjects on the college level, if industry is lacking. We can accordingly, very early in the student's college career, dissuade those of inferior capacity, who are failing to master the freshmen tasks, from attempting work to which they are not prepared to give unusual effort. In determining these eliminations at the end of the first or second semester the mental tests prove in this way of much practical assistance. Other minor practical values they have, also. To give one instance, it is judicious to present to the student who is advised to withdraw and in some cases to her parents or guardians as much evidence as possible of her unfitness to cope with the college curriculum. To relieve those who have the responsibility of recommending withdrawal of some of the onus of requesting a student of influential family to leave the institution is in itself a contribution.

Mental tests make possible a comparison of the student body with that of other colleges of like kind in a very important respect. It is of some moment to know whether a college is receiving the same proportion of able students as similar institutions, since one important element in estimating the achievement and relative standing of a college is the carrying power of its graduates, and if institutions are not receiving equally fine student material, the distinctions earned by their graduates are likely to be fewer, however fine the instruction and however ample the resources. Any administration seeking to maintain the high reputation of an institution must needs have the means of selection of students in mind, and the wise use of this new instrument is a valuable aid to success in this respect. Adequate preparation is of course an influential factor also, but thorough preparation alone will not compensate for relatively inferior ability. Indeed, no single factor contributes more to the success of a college than a student body of attested ability.

For these and other reasons it is desirable that standards for entrance to the women's colleges of liberal arts should be determined on a joint basis and that the same tests should be applied in several of these institutions. Already valuable information is at hand from the application of the Thorndike Intelligence Examina-

tion to several men's colleges of different type, to normal schools and to a group of women in a state university of the Middle West. Only by such comparative data can a thorough comprehension of the more important of the actual conditions prevailing in a particular institution be had.

Tests such as the Thorndike Intelligence Examination were originally designed for the selection of men. Some of them are admittedly ill-adapted to women, requiring such knowledge as the typical woman candidate for admission to a college is unlikely to have. Consequently, women obtain, in general, lower scores on the whole examination than men in similar institutions. A detailed survey of the differences found would be illuminating and the substitution of new tests requiring knowledge of a kind familiar to women, but unknown by the typical man, is desirable.

Intelligence tests serve a purpose still more intimately related to the successful administration of the women's college, and the realization of its aims. They make possible the classification of students on the basis of ability in the various sections of the courses required of all students. Too little attention has been paid to this desirable organization in the past. Even to-day heads of departments in the women's colleges will make the statement that a fifteen-minute test given early in a course will suffice to arrange the members of the group tested in an order of merit, which is representative of their true ability in the trait or traits measured and which remains the same in all future testings. Much evidence exists, however, as to the unreliability of such results and as to the undoubted value of grouping together those of proved similar capacity in the case of pupils in the elementary and secondary schools. While it is true that classification on the basis of similar achievement in the particular subject of study has much in its favor, nevertheless, general ability is a potent influence in progress and we ought to take it into account in classifying students where no better method is available and provided the system of assigning sections is sufficiently flexible that transfers can readily be made.

There is much waste at present in the colleges of liberal arts for women because such a system is not in operation. Inquiry along this line at Goucher college revealed a great range of differ-

ences among freshmen and notably in abilities which are fundamental to success with college work. A detailed study of the marks obtained in the reading tests in the Thorndike Intelligence Examination indicates clearly that the assignments given in such subjects as history, sociology, economics, and psychology are beyond the power of some of the students to comprehend and assimilate in the time at their disposal. There can be no doubt that, in an effort to meet the needs of the largest number, the top and bottom 20 percent are being sacrificed for the middle group of average students. Better results would follow from classification of the freshmen in required English courses on the basis of reading ability or on language ability (where all tests involving mastery of the vernacular are pooled). Moreover, the instructor's problem would be vastly simplified in having a group of similar capacity to teach.

This consigning of students to sections of like ability is in essence a phase of educational guidance. The rejection of certain candidates for entrance and the later elimination of others, are other phases of the same process, since directing students away from work for which they are unfitted is valuable for students as well as for the institution. There are other aspects of guidance in which intelligence tests can be of much assistance. The student of superior ability who receives low academic grades obviously requires different advice from the student of meager mental talents, who receives low grades. The correct location of the source or sources of failure with college work is essential to attaining efficiency, and the intelligence indices of the students make diagnosis of causes of inefficiency a more easy task. An analysis of the causes sometimes reveals conditions of which the administration was unaware. It may be that the institution is not providing an environment favorable to study. Library, laboratory or dormitory conditions may be found to be inimical to good work. Student government weakly functioning, for instance, sometimes fails to secure dormitory conditions favorable to study. On the other hand, it may be found that the individuals under consideration have remediable deficiencies, which require special attention, such as poor methods of learning, or inadequate study programs, leaving too little time for scholarly activities, or absence of scholarly ideals. Students from

small rural high schools certainly find adjustment in a large college community difficult. Often they lack training in planning out their working day, and frequently their methods of learning stand in need of correction. Lack of capacity has often been assigned as a cause for what is really to be attributed to defective training and limited past experience. The tests serve as a corrective in this connection and the official responsible for educational guidance of the students has a means of bringing pressure to bear on able students whose work has been unsatisfactory, so as to enforce the speedy acquisition of new and valuable habits.

For many reasons it would seem essential that academic grades should be as accurate as possible and should really represent the relative achievements of the students. While it is true that certain students of high intelligence may be lacking in zeal, nevertheless in the long run and in general we expect the students of superior ability to achieve most; in other words, we expect a high correlation between intelligence and college marks.

It follows that we would expect such academic subjects as select the superior women in intellect to have a disproportionate share of higher academic grades, and *vice versa.* The test results consequently can act as a valuable check on the prevailing Missouri System of marking. Investigation along this line has been made at Goucher College with a view to ascertaining the mental caliber of the students majoring in the various college subjects. So far results have been obtained for two years. The data are of course insufficient to justify us in drawing any generalization as regards Goucher College for other years. It is true, however, of the two years (the present junior and sophomore classes) that physics, mathematics, and chemistry select superior college women, while social science tends to select a mediocre and inferior group. These results are probably to be traced to local conditions, peculiar to the institution in question. Yet the fact remains that in such cases, where the poorest student majoring in physics is superior mentally to the average student majoring in social science, the applicability of the normal probability curve, even as a guide to grading, is seriously to be questioned. It would be more scientific to have grades conform to the intelligence curve typical of the group

selected by the particular subject. The plan should be generally adopted of furnishing the instructors in the various departments with the intelligence distribution for the actual students in their advanced classes of the current year, and as soon as such data are available, the intelligence distributions of their majors during a sufficiently large number of years. If it is remembered that distinctions such as Phi Beta Kappa and scholarships depend sometimes immediately and always remotely on college marks, it would seem unfair to penalize students majoring in certain fields by making the securing of a high grade much harder in some subjects than in others.

In any event those in authority should be aware of such selective influences at work. A wise administration could utilize such information to good effect. Thus, the problem of deciding for or against new requirements for majors in any department should surely be considered in this light, as well as in the light of other facts. It would seem necessary likewise that teachers should realize the mental quality of those they are training. The more thorough the knowledge of the person to be trained, the more efficient will be the instruction.

Of recent years the women's colleges have come to accept more responsibility for the guidance of students in the choice of a career. The means towards this end have been varied. Occasionally they have assumed the form of providing information through a series of lectures given by successful workers in fields open to women. Such a method has been used at Vassar and elsewhere. At Wellesley a more ambitious plan of individual consultation has been carried on, in which Miss Florence Jackson, of the Women's Industrial and Educational Union, has played the rôle of vocational adviser. The knowledge of the students' tastes and preferences so obtained has been of much value when linked with academic records of capacity. At Goucher College a beginning has been made in determining the selective effect of the various occupations from the standpoint of intelligence. It is planned to make a detailed study, not only of changes of occupational choice by the students during their four years in college, but also of subsequent success in the

occupations entered upon, and of the intelligence level of graduates entering the various fields of work.

It will be helpful, after a sufficiently large number of cases have been studied, to acquaint the student as to the ability of those in the occupation under consideration with whom she would inevitably be compared and with whom she must compete. Such knowledge, while far from constituting the whole or the major part of what needs to be known in making choice of a profession, nevertheless has real worth and may contribute to an appreciably better decision. Obviously, it needs to be supplemented in many ways, and at Goucher the improvement of methods of subjective rating of the students is being investigated together with other features in a desirable system of college records of students' abilities and achievements.

An ambitious scheme looking towards more specific vocational guidance is under way at Vassar, where a Bureau of Personnel Research is already established under the direction of the Department of Psychology. It is hoped that such a study will be made of the individual student as to make vocational guidance much more feasible.

There are other minor services that psychological tests can render in the administration of women's colleges, but they have more than justified the time, effort, and expense they involve by their improvement of methods of selecting, classifying, and grading students. They must, of course, be further improved and better adapted to women. Their results must still be carefully studied and evaluated, but there is no room for doubt that they are of great service and can afford clues of importance as to the proper action to be taken in administrative problems.

CHAPTER X

INTELLIGENCE TESTS IN COLLEGES AND UNIVERSITIES

Guy M. Whipple
Professor of Experimental Education, School of Education,
University of Michigan, Ann Arbor, Michigan

The aim of this paper is to summarize a considerable portion of the work that has been done in administering intelligence tests to college students. The material at my command is doubtless not exhaustive, but it is sufficiently complete to indicate the general situation in this field of intelligence testing.

For convenience I have cast certain portions of this summary into semi-tabular form. The table contains first of all, a list of the 29 institutions reported upon. This list begins with Brown University and concludes with Yale. It includes both private institutions, like Brown, Dartmouth, and Harvard, and state universities, like Illinois, Iowa, Michigan, Ohio, and Nebraska. It includes small institutions, like Clark, Hamline, and Reed, and large institutions like Chicago, Columbia, Harvard, and Michigan. It includes men's colleges, like Dartmouth, women's colleges, like Goucher, Sophie Newcomb, Wellesley, and Vassar, and co-educational institutions, like the majority of the list. On all these counts and in geographical distribution as well, the list may be regarded as sufficiently representative of the colleges of the United States, even if there have been important omissions.

In the second column there appear the names of the tests that have been used (mostly prior to 1921) in these institutions. The reader will note in general two types of test; first what are known as tests of general intelligence (illustrated by the Army Alpha test and the Thorndike test), and second; what may be termed tests of special aspects of intelligence (illustrated by these that appear, for instance, for the University of Chicago—number checking, constant increment, directions, etc., or for the University of Iowa or the long list for Harvard).

If we examine this column of tests more carefully, it will be evident that among the stock group tests of general intelligence, the Army Alpha test has had by far the most extended usage—it has been used, for instance, at Brown, Carnegie, Clark, Colorado Agricultural, Dartmouth, Hamline, Illinois, Michigan, Minnesota, Ohio State, Pennsylvania, Purdue, Rochester, Southern Methodist, Wyoming, and Yale, that is, in at least 16 of the 29 institutions represented. The reason for the great popularity of this particular intelligence examination is not far to seek. It was the first group intelligence test to be constructed by the joint efforts of a group of well-known psychologists; it was devised with special reference to use with adults; it has been applied in the army to more than one and three-quarters million of men (one of the really great feats of human engineering, I may add); the results have consequently reached a degree of standardization never attained by any other test; the test blanks were procurable for several months after the armistice at prices far below what other tests could be produced; the results obtained in the army far exceeded the most sanguine hopes of its makers.

Notwithstanding these many advantages, there are certain disadvantages about the Army Alpha test that are well recognized by those of us who frequently advocate its use. For one thing, it is possible for any person to buy copies of it with the keys to the answers (for example, in the book on *Army Mental Tests* by Yoakum and Yerkes), so that there would not be an insuperable obstacle to overcome for any student who wished to arm himself in advance by coaching on all five forms of the Alpha that are available. For another thing, and this is really more important, the Army Alpha examination is really somewhat too easy for the average college student. Too much of the 40 minutes used in its application is taken up with material that is perfectly simple, so that it does not act as efficiently as would a test specifically designed for a selected group of superior intelligence. Again, there is some evidence that the Army Alpha test is so phrased and constituted as to favor men over women, though this objection is not particularly serious.

TABLE I.—SUMMARY OF COLLEGES AND UNIVERSITIES SHOWING MENTAL TESTS USED AND GROUPS TESTED

Institution	Tests Used	Date	Groups Tested
1. Brown University	Army Alpha	1918	Freshmen and some others (400-500)
	Thorndike Coll. Entrance	1919	Freshmen (about 300)
	Thorndike and Special Brown Univ. test	1920	Freshmen (about 275)
2. Carnegie Institute Technology (including Margaret Morrison Carnegie School)	Army Alpha	——	Freshmen
	Trabue Completion Robinson's Range of Interest Gordon's Directions Analogies Whipple's Marble Statue Opposites	1917	114 freshmen
3. Chicago University of	Number Checking Opposites Constant Increment Directions Word Building Sentence Building Business Ingenuity Memory tests		Freshmen and other entrants
4. Clark University	Army Alpha Otis General (A and B) Otis Individual Thurstone Substitution Thurstone Reasoning Digit-Symbol Haggerty Reading Thorndike Coll. Entrance		Each freshman class, 300-400 in all
5. Colorado Agricultural College	Army Alpha (6 and 9)		500 college students and 350 prep. students
	Terman (Form A)		218 college students and 80 ex-soldiers
6. Columbia University	Thorndike Coll. Entrance	Since June 1919	Majority of freshmen 700 reported in 1920
7. Dartmouth College	Army Alpha Rating Scale Special Information Test	1920	143 freshmen of class of 1923

Institution	Tests Used	Date	Groups Tested
8. Goucher College	Thorndike Mental Alertness Thorndike Coll. Entrance Thorndike Coll. Entrance Thurstone Coll. Entrance Columbia Intelligence	1918–1919 1919–20 1920–21	98 seniors 182 freshmen 243 freshmen 150 freshmen 150 freshmen (random groups) 254 freshmen
9. Hamline	Army Alpha	1919	74 men— 145 women
10. Harvard	Yerkes-Rossy Point Scale (20 tests arranged for group exam. through lantern slides) Response to pictures Comparison of weights Memory span for digits Suggestibility Memory for unrelated sentences Comparison of terms Comprehension of questions Definition of terms Appreciation of questions Analogies Association of opposites Relational test Box test Ingenuity test Comparison of capital letters Code learning test Ball and field Geometrical construction Reproduction of diamonds Memory for designs		110 men of a class in psychology (average age of juniors and seniors 21.16) 130 women of psychology class (all seniors. Average age 22.2)
11. Illinois, University of	Army Alpha, Form 6	1919	3500 students, all classes
12. Iowa, State University of	Courtis Arithmetic (Series B) Whipple's Analogies Simpson's Opposites Completion Visualization Whipple's Information Logical Memory (The Dutch Homestead)		Freshmen 268 men 276 women
	Thorndike Coll. Entrance	1921	Freshmen

Institution	Tests Used	Date	Groups Tested
13. Michigan, University of	Thurstone, Test IV, Form 6 Army Alpha, Form A Whipple Coll. Reading, I	1921	350 probationers and 150 non-probationers
	Thurstone, Test IV, Form B Army Alpha, Form 6 Whipple Coll. Reading, II	1921	325 probationers and 50 non-probationers
	Army Alpha, Form 9 Brown Univ. Tests Whipple Coll. Reading II	1922	250 probationers and 50 non-probationers
14. Minnesota, University of	Army Alpha, Form E Army Alpha, Form 6 Analogies Opposites Trabue Completion, Scale J	1917 1919	275 freshmen 279 freshmen 200 sophomore women
15. Newcomb, H. Sophie Memorial	Color triangles Woolley Substitution Cancellation Memory (Marble Statue) Genus—Species (Woodworth-Wells) Woolley Opposites Word-Building test to half of pupils, and Ink-Blot test to the other half	1916	99 freshmen (mental tests) 32 seniors and 25 freshmen (information test)
16. Nebraska, University of	Thorndike Coll. Entrance	1921	1192 freshmen
17. Northwestern University	Trabue Completion (K & W) Hard Opposites Whipple's Information Test with substitution of 30 words, instead of marking by letters. (Brief responses required)	1916	100 freshmen
18. Ohio State University	Army Alpha, Forms 5, 6, 7, 8, 9 (Form 7 used twice) Revised Alpha	1919–20–21	5,950 (entire student body) To all new entering, 2,398 new students
19. Pennsylvania, University of	Army Alpha Witmer's Form-Board Cylinder Memory for digits Syllables, paragraph (Binet) Trabue Language test	1919	Freshmen and 186 returned soldiers 94 students in Psych. 1

Institution	Tests Used	Date	Groups Tested
20. Purdue, University of	Army Alpha		1,159 students (85% of enrollment)
21. Reed College	Standard tests on memory, association, attention, suggestion, imagination, judgment	1912–13	195 students
22. Rochester, University of	Army Alpha Otis Stanford Revision of Binet	1919–20	550 freshmen
23. Rutgers College		1920–21	freshmen
24. Southern Methodist College	Army Alpha		128 freshmen 79 sophomores 54 juniors 41 seniors
25. Texas, University of	Card Dealing Card Sorting Alphabet Sorting Mirror Drawing Spirometer		54 freshmen (boys) 52 freshmen (girls)
26. Vassar College	Woodworth-Wells Hard Opposite tests Analogies Test (Lists A and B of Woodworth and Wells) Substitution Cancellation Information Terman's Superior—Adult Tests	1917	38 seniors (with records from highest to lowest) 2 groups of 25 students
27. Washington, State University of	No statistical data	——	———
28. Wyoming, University of	Stanford Adult Test Army Alpha	1916	100 in 3 groups (freshmen, upper classmen, faculty)
	Thorndike Coll. Entrance 30 Individual Tests Will-Profile	1918–19 Summer 1919 1919 1919	143 students, all classes 60 rural school teachers 100 freshmen 145 freshmen and 104 other students 30 selected freshmen
29. Yale	Army Alpha, Forms 5 and 6		400 freshmen

Many of these objections have been met in the series of group intelligence tests prepared by Professor E. L. Thorndike for use with the freshmen at Columbia College and widely advertised as one of the standard devices for admission to that institution. These tests, as Table I shows, have been tried not only at Columbia, but also at Brown, Goucher, Iowa, Nebraska, Wyoming, also in several Normal Schools (see this Yearbook, Chapter VIII), and doubtless elsewhere. The Thorndike tests present three features that deserve mention: in the first place, their content is such that they present distinctly greater difficulty than the Army Alpha; in the second place, they are constructed by drawing material in chance lots from a large mass of previously prepared material, so that fresh examinations can be constructed for a period of years with the probability that each examination booklet will closely approximate in difficulty that of any other; in the third place, they demand a much longer time than any other intelligence tests on the market—each of the three parts of the examination takes the best part of an hour, and the total examination thus ties up a morning or an afternoon of the students' schedules. Professor Thorndike maintains that his tests show not only a man's intelligence, but also his ability to stick to a long and, at the end, somewhat distasteful task. The full Thorndike examination undoubtedly gives correlations with scholarship higher than those afforded by the Army Alpha tests, but they do not appear to exceed greatly, if at all, the correlations afforded by other special college group tests, like the Brown University tests. Thus, Professor Thorndike informs me that his entire examination affords correlations with success in the freshman year of .60; that Part I, which takes an hour, affords correlations of about .45 to .48; that Part II, which takes another hour, affords correlations of about .45; that Part III affords considerably lower correlations, but is valuable on account of its high partial correlations. He says: "I feel it my duty to add that to raise the correlation from .45 to .60 seems to me worth far more than the extra time required." Professor Colvin states that "the net correlation between the Brown University test and college marks for two terms was .60." He adds, moreover, that he could find no indication from examining data secured at Brown with the Thorndike tests that those tests showed up a 'quitter' or a man with a 'yellow

streak'." From another institution it was reported that two or three students fainted under the three-hour strain, and the faculty became indignant at this alleged imposition of hardship. Some evidence against too long an examination may be found in the recent demonstration by Hansen and Ream of Carnegie Institute of Technology, that in the 25-minute "Scrambled Alpha" test the score obtained in the first five minutes is fairly proportional with the total score (correlation 0.88), that for the first ten minutes is closely proportional (correlation 0.92) and that for the first 15 minutes virtually identical (correlation 0.96) with the total score for 25 minutes. This means that very little alteration in the standing of students would result, in that test at least, if the examination was stopped at the end of five minutes and that, to quote these writers:[1] "For practical purposes in predicting school success, the *fifteen minute test is just as satisfactory and reliable as the longer test.*" It is for this reason that I myself have preferred to devote the time for examining students to the giving of several tests of different sorts, rather than to giving a single, long, general intelligence test.

Into the merits of the several special mental tests that appear in the list this is hardly the time to go; the matter is too technical, and it is my judgment that the use of some form of general intelligence test is likely to supplant the use of tests of special aspects of mental capacity except for certain special situations. I may call attention, however, to the use of some form of reading test in one or two institutions and even to a test of arithmetical abilities, as suggesting the possible addition to intelligence testing of a limited amount of testing of certain school skills.

The third column of Table I merely indicates, where they are known, the dates when the testing has been done. That we may pass by with the comment that practically all of this work is quite recent and much of it still in the experimental stage.

The fourth column shows the groups tested at the various institutions. In a few institutions, like Illinois and Purdue, the entire student body has been tested, but in almost all the other institutions

[1] *J. of Applied Psych.*, 5: June, 1921, p. 186.

the testing has been limited to the freshmen. At Michigan, the testing has been confined to students on probation, and in part this has been an object at Clark, Columbia, Minnesota, Yale, and elsewhere. I shall return in a moment to the purposes of the testing.

In a few institutions I solicited by correspondence, information concerning the attitude of faculty and students toward the intelligence testing. Without attempting any statistical summary, it may be said that this attitude ranges from more or less scepticism through indifference to enthusiastic approval; in general, the work has been taken quite seriously and at least with open-mindedness. My experience at Michigan leads me to believe that many of the students are very keen to take mental tests; that they are anxious to learn their standing, and that they do not at all regard the testing of their mental ability in the light of an imposition, as some college administrators have feared.

To revert now to the object of the testing, it is evident that in many institutions the work is confessedly in a tentative stage or has been done purely for scientific purposes. Thus, the testing of 3500 Illinois students, as far as I know, led merely to the publication of median scores for the various classes and colleges. No attempt has been made by the administration to utilize the results in the guidance of students. Similarly with the work in several other colleges and universities. On the other hand, at Ohio State the entire student body, 5900, took the tests (and the faculty as well, I believe), and the results have been used by the deans in consultations with individual students regarding their performance in the classroom. At Michigan, the results of the tests of probationers were submitted to the administrative authorities, and have been used as one source of guidance in determining whether a given student should, or should not, be permitted to continue his university work. At Brown there exists a much more elaborate machinery for utilizing the intelligence tests. The results are made use of by a special committee whose function is to guide and counsel students in the selection of courses and in the choice of their life work.

At Columbia, intelligence ratings form one of the officially recognized means of admission to the college. To enter Columbia on

the basis of intelligence test scores, the student must have completed in an acceptable secondary school a course of four years' study. He must be able to offer three units in English, 2½ units of mathematics and at least 3 units in a foreign language. His school course must have been concerned primarily with languages, science, mathematics and history.''[2]

At Pennsylvania, students from first-class high schools whose rank is not high enough to secure a certificate may enter by either taking four examinations in subject matter or taking an examination in English and securing a certain standing in an intelligence test in which their scores are compared with those obtained from the testing of 1600 students and 200 returned soldiers.

There remain to be considered some of the typical results. I shall make no attempt here to set forth the actual statistical results concerning scores, medians, distributions, in the various tests (that is a technical matter that we may neglect for our purposes), but will confine my remarks to results that show the predictive value of the tests to their relation, in other words, to academic success.

In presenting these results, it ought to be made clear at the outset that no psychologist is foolish enough to suppose that native intelligence is the sole factor in academic success; all that is contended is that it is one factor, and probably the most important single factor, and that it is measurable by wholesale rapid methods with a reasonable degree of precision. It follows that the correlation between test scores and college marks or instructors' estimates or any other criterion of academic success will never reach perfection. On the other hand, it will always be positive and lie somewhere between 0 and plus 1.00, statistically speaking. Now, in general, a correlation above 0.30 may be regarded as of practical significance. Actual correlations between intelligence tests and academic standing seldom fail considerably to exceed this limit; they lie for the most part between 0.40 and 0.60. Let me cite a few at random: At Carnegie Institute of Technology correlations ranged in the thirties for the Thurstone Test, but reached 0.60 for a combination of five mental tests. At Brown, the correlation reached 0.60; at Chicago, correlation with instructors' estimates

[2] Quoted from T. H. Briggs, *Education*, April, 1919.

was 0.65, with the college marks was 0.43; at Yale the correlation with marks was 0.38 in one group and 0.42 in another; at Dartmouth, Army Alpha correlated 0.56 with faculty estimates of intelligence and 0.43 with scholarship, while a test termed "completion of definitions" (one of the more difficult mental tests devised for college purposes) correlated 0.55 with scholarship for 577 men, 0.54 with faculty estimates of intelligence, and 0.78 with faculty estimates of "aggressiveness," 0.75 with faculty estimates of "reliability," and 0.69 with faculty estimates of "personal impression." At Southern Methodist, Army Alpha correlated 0.52 with college grades in all four classes. These figures are sufficient to show the general outcome of mental testing so far as its relation with college marks and faculty estimates is concerned.

This matter of correlations raises a very important point that needs elucidation here. It is quite possible, in theory, and sometimes happens, in practice, that a moderate or low statistical correlation may co-exist with a high predictive value if the object is to cull out very inferior or very superior mentalities; in other words, a mental test might fail to differentiate neatly among students of medium ability and still select with considerable precision, students of poor or of excellent ability. Suppose that the primary object of testing were to locate the men who ought not to be allowed to enter the freshmen class, it would then be relatively an indifferent matter if the testing did not locate in the order in which they afterward were located by their actual classroom accomplishments the men who were admitted. From this point of view, it will be seen that numerical expressions of the degree of correlation obtained are not always of final significance; what is wanted is a list of the most inferior prospective students which will serve as a reliable prediction of their likelihood of failure later in college. A typical instance may be cited from the work at the Carnegie Institute of Technology, where, in a certain piece of experimental work, 14 women were selected by means of six mental tests as entering students whose ability was so poor as to warrant a prediction of failure; at the end of the first term every one of these 14 students was found to be in difficulty academically; some had been dropped;

some had left voluntarily, and the remainder had been placed on a two-thirds credit program. If mental tests can accomplish this much, they are of great usefulness administratively, regardless of their precision in predicting the relative standing of the students who remain.

On the other hand, a test that would 'shell out' the ones of superior ability would also have administrative significance. A suggestion that I got from conversation with a member of the faculty of a western institution (I think the University of Iowa) strikes me as worthy of mention in this connection. The suggestion was in substance; why not 'warn' the best students of their ability as well as warn the poorest students of their lack of it? More concretely, it was suggested that, after the freshmen had been examined, the top five percent should be summoned to the office of the Dean or the President and placed, as it were, "on the carpet." They would then be informed that they represented the best five percent of their class, that their innate ability was known, and that the responsibility was now definitely placed upon them to produce college records that accorded with their potential promise. The same thing could then be repeated with slight variation with the second five percent, and again with the third five percent. Here then, all that is needed is that the mental test should cull out the best mentalities, regardless of its failure to differentiate accurately among the mediocre ones. If the material of the mental test is well selected and properly pitched, there should be little difficulty on that score, because, while a good student may sometimes for one reason or another, make a poor record in a test, it is almost impossible for a poor or mediocre student to make a good record by any lucky accident. The gaining of a first-rate score may practically always be interpreted as indicative of the possession of superior mentality.

I remarked previously that no psychologist regarded intelligence as more than one important factor in academic success. A quotation from Colvin[3] will bring out this point more specifically:

[3] *Educational Monographs; the Society of College Teachers of Education,* Number X.

"In the main, there is a substantial agreement between the rating given a man in the mental tests and his academic record. However, in about fifteen percent of the cases a sufficient disagreement has been found to make it desirable to discover, if possible, the reasons for this disagreement. Personal interviews with the students whose records show such a disagreement have revealed the following facts:

I. It sometimes happens that the psychological tests fail to measure a man's real intelligence. This failure is due to various causes, most of which can be readily diagnosed as indicated below:

1. Sometimes a student tests low because of his relative unfamiliarity with the English language. This frequently happens in the case of foreign-born students, or students whose families speak in the home a foreign language. It may occasionally happen in the case of students who have had insufficient language training in the home and in the school.
2. A few students are slow, but accurate and thoughtful learners. The tests are too rapid to do such students full justice. On the other hand, the rapid but superficial learner has an undue advantage.
3. Sometimes students come from high schools where examinations are not required, and a strenuous psychological test at the beginning of their college career places them at a distinct disadvantage.
4. Emotional upsets may result in a low psychological score.
5. Lack of earnestness in taking the examination, and at times—though rarely—positive malingering, give scores far below the student's real ability.

II. The intelligence rating may be substantially correct, but other factors may weigh heavily in determining a student's success or failure in college. The most important of these are:

1. The character of the student, particularly his willingness to hold himself down to a strict mental regimen.
2. His ideals and purposes.
3. His previous educational training, including his study habits.
4. His outside distractions, including work, extra-curricular activities and social engagements.

In the light of these facts it may reasonably be concluded that psychological tests, while a valuable aid in determining a student's ability to do college work, cannot be relied upon blindly or exclusively. They must be used together with other materials as a basis for diagnosis and prognosis in connection with educational advice and direction in high school and in college."

Very similar results appeared in my own work at Michigan when some 600 students on probation were given two general intelligence tests and a college reading test of my own devising. It was my assumption that the testing would unearth a considerable number of inferior minds, but the results did not confirm the expecta-

tion. On the basis of figures obtained in the examination of army recruits it has been stated by Yoakum and Yerkes that men who secure an "A" rating in this test ought to make a first-class college record and that men who secure a "B" rating ought to be "capable of making an average record in college." Actually, 94 percent of Michigan students on probation secured either A or B in the Army Alpha test (72 percent "A," 22 percent "B,") while, of the remaining 6 percent, several were students of foreign extraction whose low score must have been in considerable measure produced by lack of ready command of English. A special problem of obvious interest is raised here, which would repay further study.

Investigation of the reports made by the probation students themselves reveals the following items as responsible, in their own opinion, for their failures (the figures are the number of times the causes assigned were reported in a total of 324 cases in the first group examined):

115	Change from high school to college conditions not fully appreciated and met
110	Health poor or handicapped by physical defect
100	High-school preparation inadequate
89	Working for self-support (2 to 7 hours per day)
60	Rooming conditions unfavorable to study
57	Never taught how to study
31	Insufficient sleep
29	Simple neglect of study
28	Illness (specific recent cases)
28	Worried about studies and prospect of failure
26	Out of school for a time
21	Military service interrupted college work
	(Miscellaneous causes less than 20 times each)

It is obvious that these categories overlap and it is true that most students report several factors, and also we must remember that nearly any one will concoct an alibi for failure if invited to do so; nevertheless, there must be some significance in this list of causes; it illustrates, in any event, that other factors than lack of intelligence operate to produce college failures, and suggests that the college has a real responsibility to arrange conditions that will be favorable to earnest work and stimulate the student to reap to the full the fruits of his potential ability.

General conclusions that may be drawn from the data gathered for this chapter are as follows:

1. Intelligence tests form a useful device in college administration, though they must be combined with other indications of the student's status to be most effective.

2. The time seems likely to arrive in the near future when the majority of college entrants will have already been given one or more intelligence examinations prior to their appearance on the college campus. There should be machinery for recording and transmitting their scores in these examinations and preferably also for translating the scores to a single (probably percentile) scale.

3. College students, as a group, take kindly to the idea of intelligence examinations. Many of them are ready to go out of their way to secure them and to discuss their rating and its bearing on their career.

4. The Army Alpha is the intelligence test thus far most widely used in the colleges, but it is evidently not the best possible test for this purpose; it is too easy and operates better to detect men who lack the minimum of intelligence necessary to do work of a passing grade than it does to differentiate among men in the higher levels of intelligence.

5. The college testing has already revealed interesting evidence of differences in the intelligence levels of groups in different parts of the country, in different institutions, in different courses and classes within the same institution.

6. There is some evidence that rating scales and other methods of appraisal for non-intellectual traits, like aggressiveness, persistence, honesty, leadership, etc., will eventually be developed that will supplant helpfully the results of intelligence tests.

SOURCES OF INFORMATION ARRANGED BY INSTITUTIONS

1. Brown University	S. S. Colvin. "Psychological Tests at Brown University," *Sch. and Soc.,* 10: 1919, 27-30. "Validity of Psychological Tests for College Entrance." *Ed. Rev.,* 60: 1920, 7-17. "Purposes and Methods of Psychological Tests in Schools and Colleges," *Ed.,* 40: 1920, 404-416. Summary of address before Society of College Teachers of Education at Atlantic City, February, 1921. Also correspondence.

2. Carnegie Institute of Technology	L. L. Thurstone. "Mental Tests for College Entrance," *J. Ed. Psych.*, 10: 1919, 129-142. Also correspondence.
3. Chicago, University of	H. D. Kitson. "Psychological Measurements of College Students," *Sch. and Soc.*, 6: 1917, 307-311. Also correspondence with Frank N. Freeman.
4. Clark University	Correspondence with Dean J. P. Porter.
5. Colorado Agricultural College	Correspondence with G. T. Avery.
6. Columbia University	T. A. Briggs, "New Columbia University Admission Plan," *Ed.*, 39: 1919, 473-480. (Very general). Dean H. E. Hawkes. "The Uses of Intelligence Tests in Colleges and Universities," *U. of Pa. Bulletin; Seventh Annual Schoolmen's Week Proc.*, 1920, pp. 260-261. Thorndike, E. L. "Intelligence Examinations for College Entrance," *J. Ed. Research*, 1: 1920, 329-337.
7. Dartmouth	Henry T. Moore. "Three Types of Psychological Rating in Use with Freshmen at Dartmouth," *Sch. and Soc.*, 13: April 2, 1921, 418-420.
8. Goucher College	A. L. Rogers. "Mental Tests as a Means of Selecting and Classifying College Students," *J. Ed. Psych.*, 11: 1920, 181-192. *Mon. Soc. Coll. Teachers Ed.*, 10: 1921, p. 55. Correspondence with Agnes L. Rogers.
9. Hamline University	G. D. Walcott. "Mental Testing at Hamline University," *Sch. and Soc.*, 10: 1920, 57-60.
10. Harvard University	R. M. Yerkes and H. E. Burtt. "Relation of Point Scale Measurements of Intelligence to Educational Performance in College Students," *Sch. and Soc.*, 5: 1917, 535-40. W. F. Dearborn. "The Measurement of Intelligence," *Psych. Bulletin*, 14: 1917, 221-224.
11. Illinois, University of	Correspondence with B. R. Buckingham. See also Yoakum and Yerkes, *Army Mental Tests.*
12. Iowa, State University of	I. King and J. M'Crory. "Freshmen Tests and the State University of Iowa," *J. Ed. Psych.*, 9: 1918, 32-46.
13. Michigan, University of	Writers' personal experience

14. Minnesota, University of	M. J. Van Wagenen. "Some Results and Inferences Derived from the Use of the Army Tests at the University of Minnesota," *J. App. Psych.*, 4: 1920, 59-72. "Has the College Student Reached his Mental Maturity when he Enters College?", *Sch. and Soc.*, 9: 1919, 663-666.
15. Newcomb, H. Sophie Memorial	Dagny Sunne. "The Relation of Class Standing to College Tests," *J. Educ. Psych.* 8: 1917, 193-211.
16. Nebraska, University of	Correspondence with Winifred Hyde.
17. Northwestern University	W. L. Uhl. "Mentality Tests for College Freshmen," *J. Ed. Psych.*, 4: 1919, 13-28.
18. Ohio State University	Ellis L. Noble and George F. Arps. "University Students' Intelligence Ratings According to the Army Alpha Tests," *Sch. and Soc.*, 11: February 21, 1920, 233-237. Correspondence with J. W. Bridges. Daily Bulletin of the University, Jan. 6, 1921.
19. Pennsylvania, University of	F. H. Reiter. "A Comparison of Test Ratings and College Grades," *Psych. Clinic*, 12: 1919, 221-229. "Educational Events," *Sch. and Soc.*, Nov. 6, 1920. George G. Chambers. "Intelligence Tests at the University of Pennsylvania," *Sch. and Soc.*, 10:1919, 548-549.
20. Purdue, University of	C. L. Roberts and G. C. Brandenburg. "The Army Intelligence Tests at Purdue University," *Sch. and Soc.*, 10: 1919, 776-778.
21. Reed College	Eleanor Rowland and Gladys Lowden. "Report of Psychological Tests at Reed College," *J. Exper. Psych.*, 1916.
22. Rochester, University of	Correspondence with L. A. Pechstein.
23. Rutgers College	Correspondence with Luther H. Martin. Results to be published in Rutger's Alumni Quarterly.
24. Southern Methodist College	H. T. Hunter. "Intelligence Tests at Southern Methodist College," *Sch. and Soc.*, 10: 1919, 437-440.
25. Texas, University of	M. Calfee. "College Freshmen and Four General Intelligence Tests," *J. Ed. Psych.*, 4: 1913, 223-231.

26. Vassar College	H. Baum and Others. "Results of Certain Standard Mental Tests as Related to the Academic Records of College Seniors," *Am. J. Psych.*, 30: 1919, 307-310. M. F. Washburn. "A Note on the Terman Superior Adult Tests as Applied to Vassar Freshmen." *Am. J. Psych.*, 30: 1919, 310.
27. Washington State College of	F. A. Thompson. "College and University Surveys," *Sch. and Soc.*, 5: 1917, 721.
28. Wyoming, University of	Correspondence with June Downey.
29. Yale	John E. Anderson. "Intelligence Tests of Yale Freshmen." *Sch. and Soc.*, 11: 1920, 417-420. Correspondence with J. E. Anderson.

CONSTITUTION OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

ARTICLE I

Name.—The name of this Society shall be "The National Society for the Study of Education."

ARTICLE II

Object.—Its purposes are to carry on the investigation and to promote the discussion of educational problems.

ARTICLE III

Membership.—SECTION 1. There shall be three classes of members—active, associate, and honorary.

SEC. 2. Any person who is desirous of promoting the purposes of this Society is eligible to active membership and shall become a member on approval of the Executive Committee.

SEC. 3. Active members shall be entitled to hold office, to vote, and to participate in discussion.

SEC. 4. Associate members shall receive the publications of the Society, and may attend its meetings, but shall not be entitled to hold office, or to vote, or to take part in the discussion.

SEC. 5. Honorary members shall be entitled to all the privileges of active members, with the exception of voting and holding office, and shall be exempt from the payment of dues.

A person may be elected to honorary membership by vote of the Society on nomination by the Executive Committee.

SEC. 6. The names of the active and honorary members shall be printed in the *Yearbook*.

SEC. 7. The annual dues for active members shall be $2.00 and for associate members $1.00. The election fee for active and for associate members shall be $1.00.

ARTICLE IV

Officers and Committees.—SECTION 1. The officers of this Society shall be a president, a vice-president, a secretary-treasurer, an executive committee, and a board of trustees.

SEC. 2. The Executive Committee shall consist of the president and four other members of the Society.

SEC. 3. The president and vice-president shall serve for a term of one year, the secretary-treasurer for a term of three years. The other members of the Executive Committee shall serve for four years, one to be elected by the Society each year.

SEC. 4. The Executive Committee shall have general charge of the work of the Society, shall appoint the secretary-treasurer, and may, at its discretion, appoint an editor of the *Yearbook*.

SEC. 5. A board of trustees consisting of three members shall be elected by the Society for a term of three years, one to be elected each year.

The Board of Trustees shall be the custodian of the property of the Society, shall have power to make contracts, and shall audit all accounts of the Society, and make an annual financial report.

SEC. 6. The method of electing officers shall be determined by the Society.

ARTICLE V

Publications.—The Society shall publish *The Yearbook of the National Society for the Study of Education* and such supplements as the Executive Committee may provide for.

ARTICLE VI

Meetings.—The Society shall hold its annual meetings at the time and place of the Department of Superintendence of the National Education Association. Other meetings may be held when authorized by the Society or by the Executive Committee.

ARTICLE VII

Amendments.—This constitution may be amended at any annual meeting by a vote of two-thirds of voting members present.

MINUTES OF THE ATLANTIC CITY MEETING OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

February 26, 1921

With President H. B. Wilson in the chair the Society tried with success the experiment of extending its meeting to two sessions, one for each part of the *Yearbook*, this in the face of most annoying disturbances during the afternoon from the hammers and cartage trucks of commercial exhibitors that surrounded the hall on the Million Dollar Pier where the meetings were held.

About 800 persons attended the first session, Saturday afternoon, 2 to 5 p. m., when the following papers were presented:

THE WORK OF THE SOCIETY'S COMMITTEE ON NEW MATERIALS OF INSTRUCTION, by the Chairman of the Committee.
F. J. Kelly, Dean of the School of Education, University of Kansas, Lawrence, Kansas.

THE PSYCHOLOGICAL APPROACH TO KINDERGARTEN SUBJECT MATTER
Nina C. Vandewalker, Specialist in Kindergarten Education, Bureau of Education, Washington, D. C.

SELECTION AND ORGANIZATION OF MATERIAL EMBODIED IN THE PRIMARY SECTION
Frances M. Berry, Kindergarten-Primary Supervisor, Baltimore, Maryland

PROJECTS FOR THE FOURTH, FIFTH, AND SIXTH GRADES
Edna Keith, Elementary Supervisor, Joliet, Illinois

THE PROJECT AND THE JUNIOR-HIGH-SCHOOL CURRICULUM
H. P. Shepherd, Principal, Junior High School, Kansas City, Kansas

PROJECT WORK FOR SUBNORMAL CHILDREN
Nellie R. Olson, Faribault, Minn.

SUGGESTED PROJECTS FROM CERTAIN EXPERIMENTAL SCHOOLS
F. D. Slutz, Principal, Morraine Park School, Dayton, Ohio

These papers were discussed by Professors Frank McMurry and W. H. Kilpatrick, of Teachers College, Columbia University, by members from the the floor and by Dean Kelly, who had introduced the program. The discussion centered about the use of the

term 'project,' and about the relative emphasis upon 'method' and upon 'curriculum' which the adoption of projects as a characteristic type of educational activity implied.

The evening session was held under more favorable conditions. The noise of the exhibitors had subsided, and the speakers could be heard by the larger audience, some 1400, who assembled at 8 o'clock for the following program:

THE WORK OF THE SOCIETY'S COMMITTEE ON SILENT READING,
By the Chairman of the Committee,
Professor Ernest Horn, State University of Iowa, Iowa City, Iowa.

THE INFLUENCE EXERTED BY THE OUTWARD FORM OF A BOOK
Florence C. Bamberger, Johns Hopkins University, Baltimore, Maryland.

ANALYSIS OF ABILITY IN READING
S. A. Courtis, Director of Instruction, Normal Training and Research, Detroit, Michigan.

THE VALUE OF SPECIFIC QUESTIONS IN SILENT READING
C. E. Germane, Dean of the School of Education, Des Moines University, Des Moines, Iowa.

INDIVIDUAL DIFFICULTIES IN SILENT READING
William S. Gray, School of Education, University of Chicago, Chicago, Illinois.

The ensuing discussion, which was opened by Dean M. E. Haggerty, of the University of Minnesota, was participated in by Professor H. O. Rugg, Mrs. Sturgis, Dean F. J. Kelly, Dean C. E. Germane, Supt. Opstadt, Miss Fanny Dunn, and others, and concluded by Professor Ernest Horn. While this discussion drifted into consideration of certain technical matters connected with the administration of schoolroom tests, the general merit of the material collected in this part of the *Yearbook* was not lost sight of; it was pointed out, for instance, by Professor Rugg that in contributions of this sort, experimental work has at last come into immediate contact with the problems of the classroom and is yielding valuable principles for the guidance of the teacher's daily work.

At the Business Meeting, held directly after the evening session, the nominating committee appointed by President Wilson submitted the following report, and upon vote of the active members present, the following were unanimously elected:

For *President,* Frederick J. Kelly, University of Kansas, Lawrence, Kansas; for *Vice-President,* Lida Lee Tall, State Normal School, Towson, Maryland; for member of the *Executive Committee,* to fill the unexpired term of Dean F. J. Kelly, J. C. Brown, President of the State Normal School, St. Cloud, Minnesota; for member of the *Executive Committee,* to serve for four years, Professor Henry W. Holmes, Harvard University, Cambridge, Massachusetts; for member of the *Board of Trustees,* to serve for three years, Professor W. W. Charters, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

The Secretary reported informally to the Society certain matters that had been under discussion by the Executive Committee earlier in the day. Thus, the Committee asked an expression of opinion on the desirability of limiting admission to one of the sessions of the Society to members of the Society. The opinion appeared to be definitely in favor of continuing the present custom of open meetings. Similarly, there seemed to be no desire to alter the plan adopted at the Chicago meeting, to which a few members had protested, of cancelling membership of those whose dues remain unpaid on January 1st. In the matter of *Yearbooks* for 1922, the Committee reported that it seemed undesirable in the present situation to devote an entire *Yearbook* to the topic proposed at the Cleveland meeting, viz.: "The Content of Courses for the Training of Teachers in Normal Schools." The Committee suggested a *Yearbook* on "The Use of Mental Tests in School Administration." Members of the Society were urged to communicate to the Secretary suggestions for other topics of educational concern that might be treated in the *Yearbooks.*

The Executive Committee endorsed the following committee to cooperate with the Division of Psychology and Anthropology of the National Research Council: Messrs. W. C. Bagley, F. W. Ballou, Ernest Horn, H. O. Rugg, and G. M. Whipple, chairman.

At both the afternoon and evening sessions the Secretary explained the aims of the Society and the conditions of membership.

GUY M. WHIPPLE,
Secretary-Treasurer.

FINANCIAL REPORT OF THE SECRETARY-TREASURER OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION,

JANUARY 13, 1921, TO DECEMBER 31, 1921, INCLUSIVE

RECEIPTS FOR 1921

Balance on hand, January 13, 1921			$ 4,702.66
From sale of *Yearbooks* by the Public School Publishing Company:			
June to December, 1920	$2,413.70		
January to June, 1921	2,697.71	$5,111.41	
Interest on savings account and bonds:			
Interest on savings to December 31, 1921	$ 23.23		
Interest on Royalty Account	35.97		
Interest on Liberty Bonds	111.21	$ 170.41	
Dues from Active and Associate Members		$3,932.17	
Total income for the year			$9,213.99
Total receipts, including initial balance			$13,916.65

EXPENDITURES FOR 1921

Publishing and Distributing Yearbooks:		
Reprinting 500 *14th Yearbook, Part II*	$ 126.00	
Reprinting 1500 *20th Yearbook, Part I*	495.30	
Reprinting 2000 *20th Yearbook, Part II*	454.50	
Printing 3000 *20th Yearbook, Part I*	1,549.10	
Printing 3000 *20th Yearbook, Part II*	1,368.66	
Typing on *20th Yearbook, Part I*	21.85	
Typing on *20th Yearbook, Part II*	20.64	
Mailing *20th Yearbook*	257.63	
Mailing *19th Yearbook* (July to January)	20.25	
Telegrams	8.13	
Premium on Fire Insurance ($5,000)	13.75	
Total cost of *Yearbooks*		$ 4,335.81
Secretary's Office:		
Secretary's salary, one year, to end of Atlantic City meeting	$ 500.00	
Secretary's expenses attending Atlantic City meeting	111.89	
Secretary's expenses attending N. E. A.—Allied Societies Conference (Cleveland)	17.82	
Bookkeeping and clerical assistance	114.16	
Stamps	42.00	
Stationery	46.25	
Checks returned	9.00	
Collection	.10	
Total for Secretary's office		$ 841.22

Paid for U. S. Treasury Certificates (10 of $100.00 denomination each)	$ 800.00	
Paid for Dominion of Canada 5½% bond, due 1929, plus accrued interest ($22.29)	1,002.04	
Total invested during 1921		$ 1,802.04
Total expenditures		$ 6,979.07

SUMMARY

Total expenditures for 1921		$ 6,979.07
Balance on hand, December 31, 1921:		
Savings Account	$ 531.53	
Checking Account	2,124.16	
Treasury Certificates	800.00	
Liberty Bonds (Cost Value)	2,386.79	
Dominion Canada Bond (Cost Value)	979.75	
Bond Interest Account	115.35	6,937.58
Total		$13,916.65

MEMBERSHIP, JANUARY 11, 1922

(Paid in advance for 1922)

Honorary members	4
Active members	446
Associate members	639
Total Membership	1,089

GUY M. WHIPPLE, *Secretary-Treasurer.*

HONORARY AND ACTIVE MEMBERS OF THE NATIONAL SOCIETY FOR THE STUDY OF EDUCATION

(Corrected to February 1, 1922)

HONORARY MEMBERS

Cook, John W., 5644 Kimbark Ave., Chicago, Ill.
DeGarmo, Charles, Cocoanut Grove, Fla.
Dewey, John, Columbia University, New York City.
Hanus, Paul H., Harvard University, Cambridge, Mass.

ACTIVE MEMBERS

Adams, Ray H., Supt. of Schools, Dearborn, Mich.
Alexander, Carter, 525 W. 120th St., New York City, N. Y.
Alexander, Thomas, Peabody College, Nashville, Tenn.
Alger, John L., Normal School, Providence, R. I.
Alleman, S. A., Supt. of Schools, Napoleonville, La.
Allen, Fiske, State Normal School, Charleston, Ill.
Allison, Samuel B., District Supt., Board of Education, Chicago, Ill.
Angell, Gertrude L., Buffalo Seminary, Bidwell Parkway, Buffalo, N. Y.
Ankeney, J. V., Asst. Prof. of Agriculture Education, Univ. of Missouri, Columbia, Mo.
Anthony, Katherine M., State Normal School, Harrisonburg, Va.
Arbaugh, W. B., Commissioner of Schools, 503 County Building, Detroit, Mich.
Ashbaugh, Ernest J., Asst. Dir. Bureau of Edu. Research, Ohio State Univ., Columbus, Ohio.
Ashley, Myron L., 7113 Normal Blvd., Chicago, Ill.
Bacon, Miss G. M., Buffalo Normal School, Buffalo, N. Y.
Badanes, Saul, P. S. No. 84, Glen More Ave., Brooklyn, N. Y.
Bagley, Wm. C., Teachers College, Columbia Univ., New York City, N. Y.
Baker, Leon, Prin. Longfellow School, Tulsa, Okla.
Baldwin, Prof. Bird T., Child Welfare Research Station, Iowa City, Ia.
Ballou, Frank W., Supt. of Public Schools, Franklin School Bldg., District of Columbia, Washington, D. C.
Bamberger, Miss Florence E., Johns Hopkins Univ., Baltimore, Md.
Banes, L. A., Prin. Mark Twain School, Tulsa, Okla.
Bardy, Joseph, 2114 N. Natrona St., Philadelphia, Pa.
Barnes, Harold, Girard College, Philadelphia, Pa.
Barnes, Percival S., Supt. of Schools, East Hartford, Conn.
Baumgardner, Nina E., Eastern S. Dak. St. Normal, Madison, S. Dak.
Bell, J. Carleton, 1032A Sterling Place, Brooklyn, N. Y.
Bender, John F., Box 625, Pittsburg, Kas.
Benedict, Ezra W., Prin. High School, Coxsackie, Greene County, N. Y.
Bennett, Mrs. V. B., Prin. Moorhead School, Pittsburgh, Pa.
Benson, C. E., Apt. 212, 509 W. 121st St., New York City, N. Y.
Benton, G. W., 100 Washington Square, New York City, N. Y.
Berry, Dr. Charles Scott, 608 Oswego Ave., Ann Arbor, Mich.

Berry, Miss Frances M., Dept. of Education, Kindergarten-Primary Supervision, Madison Ave. & Lafayette St., Baltimore, Md.
Beveridge, J. H., 508 City Hall, Omaha, Neb.
Bick, Anna, 2842-A Victor St., St. Louis, Mo.
Bird, Miss Grace E., Dept. of Psychology, R. I. College of Edu., Providence, Rhode Island.
Bjornson, J. S., Supt. of Schools, Vermillion, S. Dak.
Bobbitt, Franklin, The Univ. of Chicago, Chicago, Ill.
Bolenius, Miss Emma Miller, 46 S. Queen St., Lancaster, Pa.
Bolton, Frederick E., Univ. of Washington, Seattle, Wash.
Bowlus, Edgar S., Supt. Aberdeen Public Schools, Aberdeen, Miss.
Boyden, Wallace C., Boston Normal School, Boston, Mass.
Boyer, Chas., Supt. of Schools, Atlantic City, N. J.
Boyer, Philip A., 6320 Lawnton Ave., Philadelphia, Pa.
Bradford, Mrs. Mary D., 2603 Franklin St., Wilmington, Del.
Brady, Mary J., 3017 Lafayette Ave., St. Louis, Mo.
Bragg, Mabel C., Asst. Supt. of Schools, Newtonville, Mass.
Breed, F. S., 5476 Univ. Ave., Chicago, Ill.
Breckenridge, Miss Elizabeth, Louisville Normal School, Louisville, Ky.
Breuckner, Dr. L. J., Asst. Dean, Detroit Teachers College, Blvd. & Grand River, Detroit, Mich.
Briggs, Thos. H., Teachers College, Columbia Univ., New York City, N. Y.
Brown, Gilbert L., Marquette, Mich.
Brown, J. C., Pres. State Normal School, St. Cloud, Minn.
Brown, J. H., Prin. Irving School, Tulsa, Okla.
Brown, J. Stanley, Pres. State Normal School, DeKalb, Ill.
Buchanan, Wm. D., Dozier School, 5749 Maple Ave., St. Louis, Mo.
Buchner, Edward F., Johns Hopkins Univ., Baltimore, Md.
Buckingham, Dr. B. R., Ohio State University, Columbus, Ohio.
Buckner, Chester A., Univ. of Pittsburgh, School of Education, Pittsburgh, Pa.
Burnham, Ernest, State Normal School, Kalamazoo, Mich.
Buthod, Charles, Prin., Celia Clinton School, Tulsa, Okla.
Butterworth, Julian E., Cornell Univ., Ithaca, N. Y.
Byrd, C. E., Supt. Shreveport, La.
Byrne, Lee, 916 N. Haskell Ave., Dallas, Texas.
Calmerton, Miss Gail, 424 Old Fort Place, Fort Wayne, Ind.
Cammack, I. I., Supt. of Schools, Kansas City, Mo.
Camp, Frederic S., Supt. of Schools, 52 Hoyt St., Stamford, Conn.
Carmichael, Perry, Prin. Horace Mann School, Tulsa, Okla.
Cavan, Jordan, Asst. Professor of Educ., Butler College, Indianapolis, Ind.
Chace, Howard S., 19 Thorndike St., Beverly, Mass.
Chadsey, Charles E., Univ. of Illinois, Urbana, Ill.
Chadwick, R. D., Morgan Park School, Duluth, Minn.
Chambers, Will G., Pennsylvania State College, State College, Pa.
Chandler, J. A., Pres. William & Mary College, Williamsburg, Va.
Chapman, Prof. J. Crosby, Dept. of Education, Yale University, Graduate School, New Haven, Conn.
Chapman, Ira T., Supt. of Schools, New Brunswick, N. J.
Charters, W. W., Carnegie Inst. of Tech., Pittsburgh, Pa.
Chew, Samuel L., Supt. Dist. No. 9, 6th St. & Erie Ave., Philadelphia, Pa.
Childs, Hubert G., Indiana University, Bloomington, Ind.
Clarahan, Miss Elizabeth, Apt. 5, 1375 East 57th St., Chicago, Ill.
Clement, J. A., Northwestern Univ., 2665 Orrington Ave., Evanston, Ill.
Cleveland, Miss Elizabeth, 909 Empire Bldg., Detroit, Mich.
Cobb, Margaret V., Teachers College, Columbia University, New York City, N. Y.

Cochran, T. E., Judson College, Marion, Ala.
Cody, Frank, Supt. of Schools, 1354 Broadway, Detroit, Mich.
Coffman, Lotus D., Univ. of Minnesota, Minneapolis, Minn.
Colvin, Prof. S. S., Brown Univ., Providence, R. I.
Condon, Randall J., Supt. of Schools, Cincinnati, Ohio.
Connor, William L., Longwood High School of Commerce, E. 35th & Woodland Ave., Cleveland, Ohio.
Cook, Albert S., County Supt. of Schools, Station A., Towson, Md.
Cooke, Flora J., Francis W. Parker School, 330 Webster Ave., Chicago, Ill.
Cooley, Dr. H. C., Lawrence College, 690 Lawrence St., Appleton, Wis.
Cooley, Miss June, Dept. of Educ., Oberlin, La.
Cooper, Homer E., 425 W. 118th St., New York City, N. Y.
Coppers, A. E., Prin. Lowell School, Tulsa, Okla.
Courtis, S. A., 246 Eliot St., Detroit, Mich.
Courtright, Jocelyn, 3353 Broadway, Indianapolis, Ind.
Cox, H. S., Supt. of Schools, Covington, Ky.
Cox, P. W. L., Headmaster, The Washington School of New York, 17 East 60th St., New York City, N. Y.
Coxe, Warren W., 1466½ Pennsylvania Ave., Columbus, Ohio.
Cram, Fred D., 2222 Clay St., Cedar Falls, Iowa.
Crane, J. E., Prin. Summer Ave. School, Newark, N. J.
Crewe, Miss Amy C., Baltimore Co. Public Schools, 300 Park Ave., Baltimore, Maryland.
Crow, C. S., 32 Grant Ave., New Brunswick, N. J.
Crowley, James A., John Winthrop School, Brookford St., Dorchester, Mass.
Cubberley, Ellwood P., Leland Stanford Jr. Univ., Stanford University, Calif.
Cummins, Robert A., Asst. Prof. of Edu. & Prin. Univ. High School, Univ. of Southern California, Los Angeles, Calif.
Cunningham, Resdon J., Bozeman, Mont.
Daughters, Freeman, Prof. of Edu., State University, Missoula, Mont.
Davidson, Percy E., Leland Stanford Jr. University, Stanford Univ., Calif.
Davis, Helen, Board of Education, Jackson, Mich.
Davis, Mary C., 2231 Barclay St., Baltimore, Md.
Davis, Prof. S. B., Univ. of Pittsburgh, Pittsburgh, Pa.
Davis, S. E., President State Normal College, Dillon, Mont.
Davis, Solon P., Henry Barnard School, Hartford, Conn.
Deahl, Jasper N., Univ. of W. Va., Morgantown, W. Va.
Dean, W. A., Prin. Springdale School, Tulsa, Okla.
Dearmont, Washington S., Pres. Southeast Missouri State Teachers College, Cape Girardeau, Mo.
Deerwester, Frank, State Normal School, Pittsburg, Kan.
Dickson, Geo. A., Oceanside, Calif.
Didcoct, J. J., Peabody College for Teachers, Nashville, Tenn.
Donaldson, J. O., Supt. Armory Public School, Armory, Miss.
Doyle, Mary E., Apt. 109 Inglinde, 205 N. Tacoma St., Tacoma, Wash.
Dunkelberger, Geo. F., Waynesburg College, Waynesburg, Pa.
Dunn, Fannie W., Teachers College, Columbia University, New York City, N. Y.
Dyke, Charles B., Supt. of Schools, Millburn, N. J.
Earhart, Lida B., Teachers College, Univ. of Nebraska, Lincoln, Neb.
Eby, Frederick, Univ. of Texas, Austin, Texas.
Eby, Harvey L., 1726 Walnut St., Berkeley, Calif.
Ehler, H. F., Hughes High School, Cincinnati, O.
Elliott, C. H., Rutgers College, New Brunswick, N. J.
Elliott, E. A., 426 N. 10th St., Fredonia, Kan.
Elliott, Edward C., Chancellor, Univ. of Montana, Helena, Mont.

Ellis, A. Caswell, University of Texas, Austin, Texas.
Elson, William H., 633 S. Wabash Ave., Chicago, Ill.
Emmons, Frederick E., Supt. of Schools, Elizabeth, N. J.
Emmons, P. C., Supt. of Schools, Mishawaka, Ind.
Englehardt, N. L., Teachers College, Columbia Univ., New York City, N. Y.
Ernst, Miss L. R., Board of Education, St. Louis, Mo.
Evans, Albert W., Phillips High School, 39th & Prairie Ave., Chicago, Ill.
Evans, Edna S., 33 Walter St., Salem, Mass.
Evans, J. E., Iowa State College, Station A, Box 36, Ames, Iowa.
Evenden, E. S., Teachers College, Columbia Univ., New York City, N. Y.
Fagan, Charles, Cascade School, Seattle, Wash.
Farrington, Frederic E., Chevy Chase Schools, Washington, D. C.
Feeney, T. L., Miami Univ., Oxford, Ohio.
Felmley, David, Pres. Illinois State Normal University, Normal, Ill.
Flanders, J. K., Prof. Education, Univ. of Hawaii, Honolulu, Hawaii.
Flemming, Cecile White, 519 W. 123d St., New York City, N. Y.
Fletcher, W. H., Normal School, Oshkosh, Wis.
Foreman, W. O., Whittier School, Tulsa, Okla.
Foster, H. H., State Normal College, Ypsilanti, Mich.
Foster, J. M., Supt. of Schools, Corning, N. Y.
Frantz, A. L., Supt. of Schools, Hartford City, Ind.
French, Supt. W. C., Drumright, Okla.
Froelicher, Mitchell C., The County Day School, Kansas City, Mo.
Frost, Norman, Peabody College for Teachers, Nashville, Tenn.
Gambrill, Bessie Lee, State Normal School, Trenton, N. J.
Gard, Willis L., Ohio University, Athens, Ohio.
Garver, F. M., University, Grand Forks, Michigan.
Gates, Arthur I., Teachers College, Columbia Univ., New York City, N. Y.
Gaylord, J. S., School of Speech, Northwestern Univ., Evanston, Ill.
Germane, C. E., Dean of Education, Highland Park, Des Moines, Iowa.
Geyer, Denton L., Chicago Normal College, Chicago, Ill.
Gift, Elmer B., Manhattan, Kans.
Gilliland, A. R., Asst. Prof., Lafayette College, Easton, Pa.
Goddard, H. N., Chippewa Falls, Wis.
Goetch, E. W., Ia. St. Teachers College, Cedar Falls, Iowa.
Gosling, Thomas Warrington, Supt. of Schools, Dayton St., Madison, Wis.
Grady, C. E., El Reno, Okla.
Grape, Jacob, 333 E. 22nd St., Baltimore, Md.
Gray, Miss Olive, Asst. Supt. Hutchinson Public Schools, Hutchinson, Kan.
Gray, Wm. S., University of Chicago, School of Education, Chicago, Ill.
Greene, C. A., 1122 W. 7th St., Sedalia, Mo.
Greeson, Wm. A., Supt. of Schools, Grand Rapids, Mich.
Griffin, Margery, 4045 McPherson St., St. Louis, Mo.
Griggs, O. C., Prin. Pershing School, Tulsa, Okla.
Grote, Mamie E., 1734 E. Lafayette Ave., Baltimore, Md.
Groves, J. W., 2316 Allston Way, Berkeley, Calif.
Gruenberg, Benj. C., 418 Central Park West, New York City, N. Y.
Gullette, Albert, Prin. Holmes School, Minneapolis, Minn.
Guy, J. Freeman, Dir. of Research & Measurements, Public Schools, 718 Fulton Bldg., Pittsburgh, Pa.
Gwinn, J. M., Supt. Schools, New Orleans, La.
Hackendorf, W. M., Prin. Riverview School, Tulsa, Okla.
Haggerty, M. E., University of Minn., Minneapolis, Minn.
Hake, Miss Anna M., High School, Gettysburg, Pa.
Hall, John W., Univ. of Nevada, Dean of School of Edu., Reno, Nevada.

Hall, Madison, Supt. of City Schools, Bryan, Texas.
Hallick, Reuben P., 1154 S. Third Ave., Louisville, Ky.
Hamilton, Miss Cora M., State Normal School, Macomb, Ill.
Hamilton, Katherine, Dept. of Education, 715 Commerce Bldg., St. Paul, Minn.
Hanifan, L. J., Supt. of Schools, Welsh, W. Va.
Harris, Miss Ada VanStone, 4506 Fifth Ave., Pittsburgh, Pa.
Harvey, Nathan A., 1029 Ellis St., Ypsilanti, Mich.
Hatfield, W. R., 6030 Kenwood Ave., Chicago, Ill.
Heckert, J. W., Miami University, Oxford, Ohio.
Hendrix, Supt. H. E., Mesa, Arizona.
Henry, T. S., State Normal School, Kalamazoo, Mich.
Hernden, V. E., Prin. Lombard School, Tulsa, Okla.
Herron, Miss Lucile F., 1402 N. Alabama St., Indianapolis, Ind.
Hill, A. B., 3508 High St., Little Rock, Ark.
Hill, J. A., President West Texas State Normal School, Canyon, Texas.
Hobson, Miss Elsie G., Bryn Mawr College, Bryn Mawr, Pa.
Hoke, Kremer J., Dean of Education, College of William & Mary, Williamsburg, Va.
Holmes, Prof. Henry W., Dean, Graduate School of Education, Harvard Univ., Cambridge, Mass.
Hook, T. E., Troy, Ohio.
Horn, Ernest, 934 Kirkwood Ave., Iowa City, Iowa.
Horn, John Lewis, School of Education, Mills College, Oakland, Calif.
Hornbaker, Mrs. Esther Perry, 1710 West 104th St., Chicago, Ill.
Hosic, James F., Teachers College, Columbia University, New York City, N. Y.
Hudelson, Earl, West Virginia University, Morgantown, W. Va.
Huffaker, C. L., Supt. Mechanicsville Consol. School., Mechanicsville, Ia.
Imboden, Miss Sarah M., Super. Elem. Grades, Public Schools, Decatur, Ill.
Ingler, Francis M., Lawrence College, Appleton, Wis.
Inglis, Prof. Alexander J., Harvard University, Cambridge, Mass.
Isanogle, A. M., Thurmont, Md.
Jeffers, Fred A., Supt. of Schools, Painesdale, Mich.
Johnson, Franklin W., 60 Edgecliff Terrace, Yonkers, N. Y.
Johnson, S., P. O. Box 1073, Orlando, Florida.
Johnston, Miss Kathryn, P. O. Box 738, Helena, Mont.
Jones, Arthur J., University of Pa., Philadelphia, Pa.
Jones, Elmer E., 1115 Church St., Evanston, Ill.
Jordan, R. H., 121 Goldwin Smith Hall, Cornell Univ., Ithaca, N. Y.
Judd, Charles H., University of Chicago, Chicago, Ill.
Keeling, A. J., Prin. Jefferson School, Tulsa, Okla.
Keith, Allen P., 20 Locust St., New Bedford, Mass.
Keith, Miss Edna, Elementary Super., Joliet, Ill.
Kellcher, Josephine M., Garrison & Sheridan, St. Louis, Mo.
Kelley, Prof. Truman L., Stanford University, Palo Alto, Calif.
Kelly, F. J., University of Kansas, Lawrence, Kansas.
Kemp, W. W., Univ. of Calif., Berkeley, Calif.
Kent, Raymond A., Dean, University of Kansas, Lawrence, Kansas.
Kerr, W. H., State Normal School, Emporia, Kansas.
Kibbe, L. A., Administration, Rural Education, Washington State Normal School, Bellingham, Washington.
Kiefer, Frieda A., Room 106, Tappan Hall, University of Michigan, Ann Arbor, Michigan.
Kilpatrick, Prof. W. H., Teachers College, Columbia Univ., New York City, New York.
King, LeRoy A., University of Pennsylvania, Philadelphia, Pa.

Kirby, Thomas J., University of Iowa, Professor of Education, Iowa City, Iowa.
Kirk, John R., Pres. State Normal School, Kirksville, Mo.
Kirkham, Louisa, 1827 Orrington Ave., Evanston, Ill.
Kline, Charles W., 414 First National Bank, Waterloo, Ia.
Knapp, E. F., Super. of Grades, Syracuse, N. Y.
Knights, Maude Daisy, 1302 Harvard Blvd., Toledo, Ohio.
Koch, C. D., Dept. of Public Instr., Harrisburg, Pa.
Kohs, Samuel C., 748 Irving St., Portland, Oregon.
Kolbe, Miss Julia, Harris Bldg., Apt. 5, Norwood, Ohio.
Konold, Arthur W., Elwood, Ind.
Koonce, B. E., Prin. Lee School, Tulsa, Okla.
Koontz, Norman C., Supt. of Schools, Jamestown, N. Dak.
Koos, L. V., University of Minn., Minneapolis, Minn.
Kordsiemon, Miss A. M., Super. Special School, Quincy, Ill.
Krackowizer, Alice M., Lincoln Bldg., Box 1896, Lewiston, Mont.
Kuehn, Nita E., 2708 Clark Ave., Cleveland, Ohio.
Latham, R. H., Winston-Salem, N. C.
Layton, S. H., Supt. of Schools, Altoona, Pa.
Layton, Warren K., Room 303 Washington School, Beaubien St. and Madison Ave., Detroit, Mich.
Light, N. S., State Bd. of Edu., Hartford, Conn.
Logan, Anna E., Asst. Supt. of Schools, Denton Bldg., Cincinnati, Ohio.
Longshore, W. T., 520 West 40th Street, Kansas City, Mo.
Lord, Dr. L. C., State Normal School, Charleston, Ill.
Lowry, Charles D., 1643 Kenilworth Ave., Chicago, Ill.
Luckey, G. W. A., Bureau of Education, Washington, D. C.
Lukens, Herman T., 330 Webster Ave., Chicago, Ill.
McAllister, Cloyd N., Berea, Ky.
McCabe, E. S., Supt. of Schools, Copan, Okla.
McCarthy, J. C., 789 Orange St., New Haven, Conn.
McDonald, Robert A. F., Bates College, Lewiston, Maine.
McFarland, Geo. A., Pres. State Normal School, Williston, N. Dak.
McKenny, Charles, Pres. State Normal, Ypsilanti, Mich.
McLeod, J. A., Principal Lincoln School, Tulsa, Okla.
McMullin, Walter G., 17th & Pine Sts., Philadelphia, Pa.
McMurry, Frank M., 9 Hillside Drive, Yonkers, N. Y.
MacMillan, D. P., Suite 600, 460 S. State St., Chicago, Ill.
Magee, M., Prin. Osage School, Tulsa, Okla.
Maguire, Anna, 34 Sherman St., Roxbury, Mass.
Manahan, J. L., Univ. of Virginia, University, Va.
Mangun, Vernon L., Pres. Forestry S. N. S., Bottineau, N. Dak.
Marrs, S. M. N., Chief High School Super., Austin, Texas.
Marsh, Arthur L., Washington Education Association, 707 Lowman Bldg., Seattle, Wash.
Martin, J. W., 4428 W. Belle Place, St. Louis, Mo.
Mayberry, Lawrence, W., Wichita, Kan.
Maynard, M. M., Dept. of Edu., Monmouth College, Monmouth, Ill.
Mead, Alice M., Barry Public School, 59th & Race Sts., West Philadelphia, Pa.
Mead, Prof. Arthur R., Ohio Wesleyan University, Delaware, Ohio.
Mead, Cyrus D., Univ. of California, Berkeley, Calif.
Meek, Charles S., Supt. of Schools, Toledo, Ohio.
Melcher, George, Library Bldg., 9th & Locust Sts., Kansas City, Mo.
Mellyn, Mary C., care of School Com., Mason St., Boston, Mass.
Mendenhall, Edgar, Dept. of Rural Educ., State Manual Training Normal School, Pittsburg, Kan.

Meriam, J. L., 4411 Hermosa Way, San Diego, Calif.
Miller, C. S., Supt. of Schools, Latrobe, Pa.
Miller, Irving E., State Normal School, Bellingham, Wash.
Miller, J. A., Prin. Washington School, Tulsa, Okla.
Miller, Ruby A., 604 N. Franklin St., Wilmington, Del.
Miller, W. S., College of Education, University of Minnesota, Minneapolis, Minn.
Mills, Harriette M., New York University, Washington Square, New York City.
Minnich, H. C., State Normal School, Oxford, Ohio.
Monroe, Walter S., Univ. of Illinois, Urbana, Ill.
Moore, Clyde B., Dir. University Demonstration Schools, Univ. of Pittsburgh, Pittsburgh, Pa.
Morey, Chas. C., Prin. School No. 59, Buffalo, N. Y.
Morris, Fannie M., Plummer School, East Boston, Mass.
Morris, J. V. L., Dept. of Education and Psychology, Pennsylvania College for Women, Pittsburgh, Pa.
Morrison, J. Cayce, Specialist in Educational Measurements, Univ. of the State of New York, Albany, N. Y.
Morrison, H. C., The University of Chicago, Chicago, Ill.
Morse, Lucia B., Downer's Grove, Ill.
Morton, R. L., 18 Maplewood Drive, Athens, Ohio.
Morton, Wm. M., 330 East 22nd St., Chicago, Ill.
Muerman, J. C., Bureau of Education, Washington, D. C.
Munson, O. F., Roswell, New Mexico.
Murphy, Frederick J., 574 Eighth St., South Boston, Mass.
Neuharth, John J., Athol, S. Dak.
Neverman, P. F., Supt. of Schools, Marinette, Wis.
Newlon, Jesse H., Supt. of City Schools, Denver, Colo.
Newman, Hugo, 538 W. 150th St., New York City.
Nichols, C. A., Southern Meth. Univ., Dallas, Texas.
Nifenecker, Eugene A., 390 Wadsworth Ave., New York City, N. Y.
Nixon, Carl, Prin. Central Grade, Tulsa, Okla.
Noble, Stuart G., Millsaps College, Jackson, Miss.
Oberholtzer, E. E., Supt. of Schools, Tulsa, Okla.
Olson, Nellie R., 314 7th Ave., Faribault, Minn.
O'Neal, Joseph A. F., John Winthrop School, Boston, Mass.
Orman, Clarence, Prin. Kendall School, Tulsa, Okla.
Osburn, W. J., Director of Educational Measurements, Madison, Wis.
O'Shea, M. V., Univ. of Wisconsin, Madison, Wis.
Ottermann, Charles, 3301 Observatory Ave., Cincinnati, Ohio.
Palmer, Louisa, School of Childhood, Univ. of Pittsburgh, Pittsburgh, Pa.
Parker, Samuel C., Univ. of Chicago, Chicago, Ill.
Partridge, Mrs. Clara Martin, 2413 Milvia St., Berkeley, Calif.
Paul, J. B., Ia. State Teachers College, Cedar Falls, Ia.
Pauly, Frank R., Supt. of Schools, Glenpool, Okla.
Pearson, M. E., Library Bldg., Kansas City, Kansas.
Pechstein, L. A., University of Rochester, Rochester, N. Y.
Pendleton, Charles S., Asst. Supt. of Schools, Winnetka, Ill.
Penfold, Arthur, 332 Beard Ave., Buffalo, N. Y.
Perkins, Geo. W., 342 Madison Ave., New York City, N. Y.
Perrine, C. H., Parker High School, Chicago, Ill.
Phelan, W. W., 536 Chautauqua St., Norman, Okla.
Philhower, Chas. A., Westfield, N. J.
Phillips, Edna K., N. Y. Training Sch. for Teachers, 220 West 120th St., New York City.

Pillsbury, W. H., Board of Education, Buffalo, N. Y.
Pintner, Rudolph, Teachers College, Columbia University, New York, N. Y.
Piper, Ernest E., 2326 Woodward Ave., Detroit, Mich.
Pittenger, B. F., University of Texas, Austin, Texas.
Pittman, M. S., Ypsilanti, Mich.
Porter, David C., Bridgeton, N. J.
Power, Leonard, 312 DeQueen Blvd., Port Arthur, Tex.
Powers, S. R., Univ. of Ark., College of Edu., Fayetteville, Ark.
Pratt, O. C., Supt. of Schools, Administration Building, Spokane, Wash.
Price, E. D., Enid, Okla.
Prunty, M. C., Prin. Central High School, Tulsa, Okla.
Pryor, Hugh C., Northern Normal & Indus. School, Aberdeen, S. Dak.
Race, Henrietta V., Board of Education, Youngstown, Ohio.
Rader, L. W., Asst. Supt. 911 Locus Ave., St. Louis, Mo.
Rall, E. E., Pres. North-Western College, Naperville, Ill.
Rankin, P. T., Detroit Teachers College, Detroit, Mich.
Rapeer, Louis W., Research University, 20 Jackson Place, Washington, D. C.
Rawland, Albert Lindsay, Dept. of Public Instruction, Harrisburg, Pa.
Ream, C. H., Supt. of Schools, Clear Lake, Iowa.
Reeve, Wm. D., Univ. High School, Univ. of Minnesota, Minneapolis, Minn.
Reid, O. L., 20 W. Wood St., Youngstown, Ohio.
Rhoton, Prof. A. L., Pa. State College, State College, Pa.
Rice, Clarence I., 934 Barnett St., Kansas City, Kansas.
Richards, J. P., Prin. Sequoyah School, Tulsa, Okla.
Rinsland, Henry D., Dir. Dept. Research & Guidance, Ardmore City Schools, Ardmore, Okla.
Robbins, Chas. L., Prof. of Education, 1049 Woodlawn, S. U. of I., Iowa City, Iowa.
Rogers, Agnes L., Goucher College, Baltimore, Md.
Root, Chas. C., State Normal School, Buffalo, N. Y.
Rosier, Joseph, Pres. Normal School, Fairmont, W. Va.
Rounds, C. R., Supt. of Schools, Ft. Thomas, Ky.
Roy, Victor Leander, Natchitoches, La.
Ruedniger, Geo., Washington Univ., Washington, D. C.
Rugg, Harold Ordway, Lincoln School, 646 Park Ave., New York City, N. Y.
Rynearson, Edw., Prin. Fifth Ave., High School, Pittsburgh, Pa.
Sachs, Dr. Julius, Columbia University, 225 W. 86th St., New York City.
Sailer, T. H. P., Englewood, N. J.
Sanger, Wm. T., 312 Lyric Bldg., Richmond, Va.
Schumacher, Very Rev. Matthew, Pres. of St. Edwards College, Austin, Texas.
Schwiering, O. C., Rock Springs Public Schools, Rock Springs, Wyoming.
Scott, Z. E., Supt. of Schools, Louisville, Ky.
Searles, Clair K., Room 106, Tappan Hall, Univ. of Mich., Ann Arbor, Mich.
Selke, George, State Dept. of Edu., St. Paul, Minn.
Sexson, J. A., Sterling, Colo.
Shankland, Sherwood D., Andrews Institute, Willoughby, Ohio.
Shotwell, Fred C., Prin., 121 Westervelt Ave., Plainfield, N. J.
Sickel, Edward, Prin. of School 27, Fayette & Chester Sts., Baltimore, Md.
Sies, Raymond W., Univ. of Kentucky, Lexington, Ky.
Sister Marguerita, St. Ursula's Academy, Toledo, Ohio.
Slater, C. P., Supt. of Schools, West Liberty, Ia.
Slauson, Herbert M., 433 So. Fifth Ave., Ann Arbor, Mich.
Slutz, Frank D., Moraine Park School, Box 782, Dayton, Ohio.
Smart, Frank L., Supt. of Schools, Davenport, Iowa.
Smiley, Wm. H., Asst. Supt. Schools, Commonwealth Bldg., Denver, Colo.

Smith, Aubrey G., Asst. City Supt. of Schools, Medford, Oregon.
Smith, Eleanor L., Rockville, Md.
Smith, Eugene R., Headmaster, Park Sch., Baltimore, Md.
Smith, H. L., Dean, School of Edu., University of Indiana, Bloomington, Ind.
Smith, Jesse L., Supt. of Schools, 141 So. 2nd St., Highland Park, Ill.
Smith, Leon O., Asst. Supt. Schools, Bd. of Edu., Omaha, Neb.
Smith, Mrs. Margaret M., Prin., Maria Mitchell Sch., Denver, Colo.
Smith, Marion C., 1832 Hopkins St., Berkeley, Calif.
Smith, Wm. F., Prin. Elwood High School, 1920 North A St., Elwood, Ind.
Smoot, Lucy J., 4011 Baltimore, Kansas City, Mo.
Smyth, Miss E. A., 1217 Madison Ave., Baltimore, Md.
Snedden, David, Teachers College, Columbia University, New York City.
Soltes, Mordecai, Dir. Jewish Extension Center, The Bureau of Jewish Education, 31 W. 110th St., New York City, N. Y.
Spencer, W. L., Supervisor of Secondary Education, Montgomery, Ala.
Sprague, H. A., 121 Summit Ave., Summit, N. J.
Stanley, G. A., Central School, Tacoma, Washington.
Starch, Daniel, 89 Trowbridge, Cambridge, Mass.
Stark, William E., Supr. Prin. of Schools, Hackensack, N. J.
Steele, Roxana A., Dir. Training School, Bloomsburg, Pa.
Stilwell, William Earle, Headmaster, University School, Cincinnati, Ohio.
Stockinger, W. A., Supt. of Schools, Noblesville, Ind.
Stone, Chas. R., Supt. of Schools, 1518 Maple St., Homestead, Pa.
Stone, Cliff W., 603 Linden Ave., Pullman, Wash.
Stoutemyer, J. Howard, 1930 8th Ave., Kearney, Neb.
Strayer, G. D., Teachers College, New York City.
Strecker, John K., Baylor University Library, Waco, Texas.
Strong, Edward K., Jr., Carnegie Inst. of Tech., Pittsburgh, Pa.
Stuart, Josephine B., 87 Mill St., New Bedford, Mass.
Study, Harry P., 309 R. St., Atchison, Kan.
Sutherland, Anne, Walnut Hills High School, Cincinnati, Ohio.
Sutherland, Dr. A. H., Board of Education, Security Bldg., Los Angeles, Calif.
Suzzallo, Henry, Univ. of Washington, Seattle, Wash.
Tall, Miss Lida Lee, Princ. of Maryland State Normal School, Towson, Md.
Tanger, Landis, Supt. of Schools, Homestead, Pa.
Taylor, J. S., Dist. Supt. of Schools, 2275 Loring Place, Bronx, New York City, N. Y.
Terry, Paul W., University of Chicago, School of Education, Chicago, Ill.
Tew, Derwood J., Super. Grammar Grades, Camden, N. J.
Theisen, W. W., Dir. Board of Edu., Cleveland, Ohio.
Thomas, A. D., Supt. of Schools, Hazleton, Pa.
Thompson, Clement O., Prof. of Education, Hanover College, Hanover, Ind.
Thompson, Frank E., University of Colorado, Boulder, Colo.
Thorndike, E. L., Columbia University, New York City.
Thurber, C. H., Editor, Ginn & Co., Boston, Mass.
Tibbets, Anna, Fargo College, Fargo, N. Dak.
Tidwell, R. E., Asst. Supt. Dept. of Edu., Montgomery, Ala.
Tilley, James M., 6½ Walnut St., Terre Haute, Ind.
Tippett, Jas. S., Peabody College, Nashville, Tenn.
Tirapegui, Luis A., Casilla 604, Santiago, Chile.
Toaz, Robert K., Huntington, Long Island, N. Y.
Towne, Chas. F., Lasell Seminary, Auburndale, Mass.
Trabue, M. R., Teachers College, Columbia University, New York City.
Traner, F. W., University of Nevada, Reno, Nev.

Truesdell, Benj. W., 412 N. Emporia Ave., Wichita, Kan.
Trumper, May, Helena, Mont.
Updegraff, Harlan, University of Penna., Philadelphia, Pa.
Vandewalker, Miss Nina C., Bureau of Education, Washington, D. C.
Van Sickle, Edith S., Dover, Del.
Van 't Roer, Cecelia, 115 W. Bush St., Kalamazoo, Mich.
Verplanck, Fred A., So. Manchester, Conn.
Vincent, H. D., Prin., Pub. Sch. No. 3, Cor. 5th Ave. & J. St., Troy, New York.
Volker, Wm., Main, 2nd and 3rd Sts., Kansas City, Mo.
Wagg, Alvin P., Oliver Wendell Holmes Sch., Dorchester Centre, Mass.
Wahlert, H. H., Asst. Supt., North Bergen, N. J.
Waldo, Dwight B., Western State Normal School, Kalamazoo, Mich.
Walker, D. E., 1011 Washington St., Evanston, Ill.
Walker, E. T., Prof. Edu., Junior College, El Paso Public Schools, 3222 Aurora St., El Paso, Texas.
Walters, R. J., Rocky Ford, Colo.
Warner, W. W., 621 S. Jefferson St., Saginaw, Mich.
Washburne, Supt. Carleton W., Winnetka, Ill.
Watkins, C. B., Prin. Emerson Sch., Tulsa, Okla.
Watts, Rowland, 3315 Powhatan Ave., Baltimore, Md.
Weathersby, Wm. H., Prof. of Edu., Mississippi College, Clinton, Miss.
Weber, A. W., 2207 E. 85th St., Cleveland, Ohio.
Weber, S. E., Supt. of Schools, Scranton, Pa.
Weglein, David E., 2400 Linden Ave., Western High School, Baltimore, Md.
Welsh, Wm. Henry, 17th & Pine Sts., Grant Building, Philadelphia, Pa.
West, Henry S., Supt. of Public Instruction, School Administration Bldg., Baltimore, Maryland.
West, Paul V., Asst. Prof. Education, College of Education, University of Wisconsin, Madison, Wis.
Whipple, G. M., University of Michigan, Ann Arbor, Mich.
White, L. A., Supt., Minot, N. D.
Whitney, Prof. A. S., Univ. of Michigan, Ann Arbor, Mich.
Whitney, F. L., 715 University Ave., Minneapolis, Minn.
Wickey, Miss Rose, 3031 Paseo, Kansas City, Mo.
Wilber, Flora, Normal School, Fort Wayne, Ind.
Wilcox, Walter H., Prin. Monroe School, 3701 S. Broadway, St. Louis, Mo.
Wilde, Arthur H., 125 Fair Oaks Park, Needham, Mass.
Williams, E. I. F., 42 Circular St., Tiffin, Ohio.
Willing, Matthew H., Prin. H. School, Springfield, Ill.
Wills, Benj. G., Estancia, N. Mexico.
Wilson, Mrs. L. L. W., South H. S. for Girls, Philadelphia, Pa.
Witham, Ernest C., Supt. of Schools, Southington, Conn.
Wood, O. A., 4213 E. 58th St., Kansas City, Mo.
Woody, Clifford, University of Mich., Ann Arbor, Mich.
Yocum, A. D., Univ. of Penn., Philadelphia, Pa.
Young, Leonard, Prin. Central High School, Duluth, Minn.
Ziegler, J. W., 1006 Arch St., care of John C. Winston Co., Philadelphia, Pa.
Zimmers, P. J., Supt. of Schools, Boise, Idaho.
Zirbes, Laura, Investigator of Reading, Lincoln School, New York City, N. Y.

Information Concerning the National Society for the Study of Education

1. PURPOSE. The purpose of the National Society is to promote the investigation and discussion of educational questions. To this end it holds an annual meeting and publishes a series of Yearbooks.

2. ELIGIBILITY TO MEMBERSHIP. Any person who is interested in receiving its publications may become a member upon application to the Secretary and subsequent approval by the Executive Committee. Membership may not be had by libraries or by institutions.

3. PERIOD OF MEMBERSHIP. Applicants for membership may not date their entrance back of the current calendar year, and all memberships terminate automatically on December 31st, unless the dues for the ensuing year are paid as indicated in Item 6.

4. CLASSES OF MEMBERS. Application may be made for either active or associate membership. Active members pay two dollars dues annually, receive two copies of each publication, are entitled to vote, to hold office and to participate in discussion. Associate members pay one dollar dues annually, receive one copy of each publication, may attend the meetings of the Society, but may not vote, hold office or participate in discussion. The names of active members only are printed in Part I of each Yearbook. There were in 1921 about 400 active and 800 associate members.

5. ELECTION FEE. New active and new associate members are required the first year to pay, in addition to the dues, an election fee of one dollar.

6. PAYMENT OF DUES. Statements of dues are rendered in October for the following calendar year. By vote of the Society at the 1919 meeting, "any member so notified whose dues remain unpaid on January 1st, thereby loses his membership and can be reinstated only by paying the election fee of one dollar required of new members."

7. DISTRIBUTION OF YEARBOOKS TO MEMBERS. The Yearbooks, ready each February, will be mailed only to members whose dues for that year have been paid. Members who desire Yearbooks prior to the current year must purchase them directly from the publishers (see Item 8).

8. COMMERCIAL SALES. The distribution of all Yearbooks prior to the current year and also of those of the current year not regularly mailed to members in exchange for their dues is in the hands of the publishers, not of the secretary. For such commercial sales, communicate directly with the Public School Publishing Company, Bloomington, Illinois, who will gladly send a price list covering all the publications of this Society and of its predecessor, the National Herbart Society.

9. YEARBOOKS. The Yearbooks are issued in parts (usually two) every February. They comprise from 250 to 500 pages annually. Unusual effort has been made to make them on the one hand of immediate practical value and on the other hand representative of sound scholarship and scientific investigation. Many of them are the fruit of cooperative work by committees of the Society.

10. MEETINGS. The annual meeting, at which the Yearbooks are discussed, is held in February at the same time and place as the meeting of the Department of Superintendence of the National Education Association.

Applications for membership will be handled promptly at any time on receipt of name and address, together with check for the appropriate amount ($3.00 for new active membership, $2.00 for new associate membership).

GUY M. WHIPPLE, Secretary-Treasurer.

University of Michigan,
Ann Arbor, Michigan.